

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)



PCT / IB 00 / 0 1 0 2 6
12.09.00



INVESTOR IN PEOPLE

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

1800/1026

REC'D 13 SEP 2000

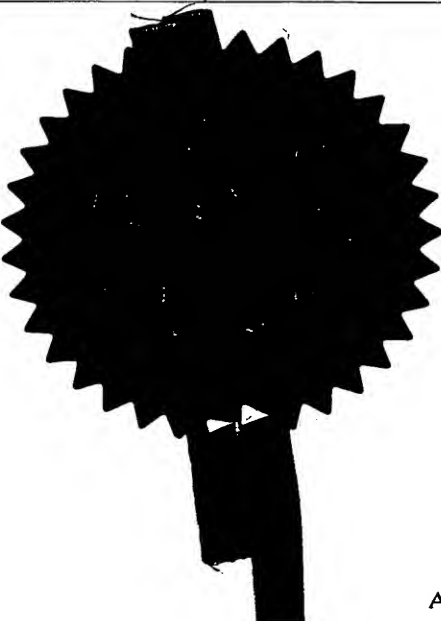
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



R. Mahoney

Signed

Dated

19 JUL 2000

THIS PAGE BLANK (USPTO)

The
Patent
Office

1/77
15 JUL 99 E462090-1 D00019
P01/7700 0.00 - 9916529.2

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference **P022260GB**

2. Patent application number
(The Patent Office will fill in this part) **9916529.2**

3. Full name, address and postcode of the or of each applicant (underline all surnames)
**CHIRON SpA
Via Fiorentina 1
53100 Siena
ITALY**

Patents ADP number (if you know it)

07157811001

If the applicant is a corporate body, give the country/state of its incorporation

ITALY

4. Title of the invention **ANTIGENIC PEPTIDES**

5. Name of your agent (if you have one) **Carpmaels & Ransford**

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

**43 Bloomsbury Square
London
WC1A 2RA**

Patents ADP number (if you know it)

83001

If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an applicant, or
- c) any named applicant is a corporate body

Yes

See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form -

Description 156

Claim(s) 1

Abstract -

Drawing(s) -

10. If you are also filing any of the following, state how many against each item.

Priority documents -

Translations of priority documents -

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*) -

Request for preliminary examination and search (*Patents Form 9/77*) -

Request for substantive examination (*Patents Form 10/77*) -

Any other documents (please specify) -

11. I/We request the grant of a patent on the basis of this application.

Signature

Date

Carmichael & Ransford
Carmichael & Ransford

14th July 1999

12. Name and daytime telephone number of person to contact in the United Kingdom

Huw G Hallybone

0171 242 8692

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
- If you have answered 'Yes' Patents Form 7/77 will need to be filed.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

ANTIGENIC PEPTIDES

This invention relates to antigenic peptide sequences from the bacteria *Neisseria meningitidis*.

BACKGROUND

Neisseria meningitidis is a non-motile, gram negative diplococci that are pathogenic in humans.

- 5 Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries.
- 10 The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Meningococcus B remains a problem, however. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. One approach to a menB vaccine uses mixtures of outer membrane proteins (OMPs) To overcome the
- 15 antigenic variability, multivalent vaccines containing up to nine different porins have been constructed [eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28]. Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability [eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both
- 20 ~~surface exposed and generate bactericidal antibodies capable of killing homologous and~~ heterologous strains. *Vaccine* 14(1):49-53].

THE INVENTION

The invention provides fragments of the proteins disclosed in International patent application PCT/IB99/00103 [Annex 1], wherein the fragments comprise at least one antigenic determinant.

- 25 Thus, if the length of any particular protein sequence disclosed in PCT/IB99/00103 is x amino acids (see Table II), the present invention provides fragments of at most $x-1$ amino acids of that protein. The fragment may be shorter than this (eg. $x-2$, $x-3$, $x-4$, ...), and is preferably 100 amino

acids or less (eg. 90 amino acids, 80 amino acids *etc.*). The fragment may be as short as 3 amino acids, but is preferably longer (eg. up to 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 35, 40, 50, 75, or 100 amino acids).

5 Preferred fragments comprise the meningococcal peptide sequences disclosed in Table I, or sub-sequences thereof. The fragments may be longer than those given in Table I *eg.* where a fragment in Table I runs from amino acid residue p to residue q of a protein, the invention also relates to fragments from residue $(p-1)$, $(p-2)$, or $(p-3)$ to residue $(q+1)$, $(q+2)$, or $(q+3)$.

10 The invention also provides polypeptides that are homologous (ie. have sequence identity) to these fragments. Depending on the particular fragment, the degree of sequence identity is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more). These homologous polypeptides include mutants and allelic variants of the fragments. Identity between the two sequences is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

15 The invention also provides proteins comprising one or more of the above-defined fragments.

The invention is subject to the proviso that it does not include within its scope proteins comprising any of the 45 protein sequences disclosed in PCT/IB99/00103 (*ie.* the even SEQ IDs: 2, 4, 6, 8, 10, ..., 86, 88, 90 of Annex 1).

20 The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, C-terminal and/or N-terminal fusions *etc.*). They are preferably prepared in substantially pure form (*ie.* substantially free from other Neisserial or host cell proteins). Short proteins are preferably produced using chemical peptide synthesis.

25 According to a further aspect, the invention provides antibodies which recognise the fragments of the invention, with the proviso that the invention does not include within its scope antibodies which recognise one of 45 complete protein sequences in Annex I. The antibodies may be polyclonal or, preferably, monoclonal, and may be produced by any suitable means.

The invention also provides proteins comprising peptide sequences recognised by these antibodies. These peptide sequences will, of course, include fragments of the meningococcal proteins in Annex I, but will also include peptides that mimic the antigenic structure of the meningococcal peptides when bound to immunoglobulin.

- 5 According to a further aspect, the invention provides nucleic acid encoding the fragments and proteins of the invention, with the proviso that the invention does not include within its scope nucleic acid encoding one of the 45 protein sequences in Annex 1.

10 In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to these sequences. Furthermore, the invention provides nucleic acid which can hybridise to these sequences, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

15 Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*). In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

20 According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

25 The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines or as immunogenic compositions) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised

against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*) but are preferably *N.meningitidis*, especially strain A or strain B.

5 The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

10 A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

15 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (~~eg. to utilise the disclosed sequences for vaccination or diagnostic purposes~~)
20 follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

25 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid*

- Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).
- 5
- 10 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference.

Definitions

- A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.
- 15

The term "comprising" means "including" as well as "consisting" *eg.* a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- 20 The term "antigenic determinant" includes B-cell epitopes and T-cell epitopes.

- The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a meningococcal sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.
- 25

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain
5 origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

Expression systems

10 The meningococcal nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

15 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at
20 the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

25 Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide

useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

5 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of*
10 *the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally,
15 some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the
20 recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence
25 fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a
30 foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 5 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) 10 "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as 15 an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or 20 polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicon systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria 25 shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in 30 liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon

will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These
5 include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of
10 replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end
15 of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide
20 particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

25 DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also
30 appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect

origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified

by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by

gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants.

The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host.

The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for
5 Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

- 10 Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding
15 additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant
20 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

- A heterologous coding sequence may be for any protein relating to the present invention. The
25 sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may
30 also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested.

Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

- 5 Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.
- 10 The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of
15 small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.
- 20 The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength

reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.
- 25 All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the
30 genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*,

Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersion, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an

operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g-laotamase* (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a

hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.*

ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a

prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655];

Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology I*:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH),
5 hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription
10 activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*,
15 *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*,
[Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979)
20 "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be
25 directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian,
30 baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of

heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the

polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci. USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg.* Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers

may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example,
5 the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

10 Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol.*
15 *Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guillermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA*
20 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali
25 cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; *Hansenula*]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135;
30 *Kluyveromyces*]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; *Pichia*]; [Hinnen *et al.* (1978) *Proc.*

Natl. Acad. Sci. USA 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Antibodies

5 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies,
10 humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying meningococcal proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by
15 conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the
20 mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo*
25 immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature*
30 (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as

described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing
5 membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind
10 specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly
15 ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a
20 monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may
~~serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as~~
25 antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

- 5 The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the
10 subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

- For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50
15 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

- A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any
20 pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity.

-
- Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in
25 the art.

- Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences
30 (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject.
10 The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or
15 transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

20 Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid
25 copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

- 5 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of
- 10 the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.
- 15 The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with
- 20 other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

- 25 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo*
- 30 can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J*

Virol 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. ~~Alternatively, administration of DNA linked to killed adenovirus as described in~~ Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive

nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses, Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc*
10 *Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci*
15 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example
20 ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example
ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for
25 example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus,
30 for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic

acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial
5 No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci*
10 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell
15 targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex
20 beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796,
25 WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid,
30 insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or

ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hypodermic sprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

- 5 Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

10 Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A.Polypeptides

- One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR);
15 transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from
20 other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, etc.

- 25 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-,

or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting

materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem*

261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

- 10 Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys*
15 *Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

- 20 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously,
25 etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and

therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

- 5 The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

- Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when
10 combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

- Meningococcal antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-meningococcal antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive
15 diagnostics methods. Antibodies to meningococcal proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation.
20 Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

- Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are
25 constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, etc.) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{formamide}) - 600/n - 1.5(\% \text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

10 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not
15 completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

20 In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to
25 start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to “hybridize” with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the meningococcal nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native meningococcal sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the meningococcal sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional meningococcal sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a meningococcal sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a meningococcal sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

5 The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired meningococcal sequence.

20 A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the meningococcal sequence (or its complement).

25 Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

EXAMPLES OF PREFERRED FRAGMENTS

The protein sequences disclosed in PCT/IB99/00103 have been subjected to computer analysis to predict antigenic peptide fragments within the full-length proteins. Three algorithms have been used in this analysis:

- 5 • **AMPHI** This program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9] and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).
- **ANTIGENIC INDEX** as disclosed by Jameson & Wolf (1988) The antigenic index: a novel
0 algorithm for predicting antigenic determinants. *CABIOS*, 4:181:186
- **HYDROPHILICITY** as disclosed by Hopp & Woods (1981) Prediction of protein antigenic determinants from amino acid sequences. *PNAS*, 78:3824-3828

15 Table I indicates preferred fragments of the proteins disclosed in Annex I. The three algorithms often identify the same fragments (*eg.* ORF100 – the fragment from residue 98 to residue 109, and the fragments from residue 111 to residue 121). Such multiply-identified fragments are particularly preferred. The algorithms often identify overlapping fragments (*eg.* ORF100 – AMPHI identifies residues 143-152, and Antigenic Index identified residues 148-157). The invention explicitly includes fragments resulting from a combination of these overlapping fragments (*eg.* the fragment from residue 143 to residue 157, in the case of ORF100). Fragments
20 separated by a single amino acid are also often identified (*eg.* ORF48-1 hydrophilicity 334-342 and 344-349). The invention also includes fragments spanning the two extremes of such “adjacent” fragments (*eg.* 334-349 for ORF48-1).

TABLE I – 1769 fragments of the proteins disclosed in Annex I.

Key:

- 25 – SEQ ID 1 of the present application is amino acids 6 to 14 of ORF38-1 disclosed in Annex I, *etc.*

SEQ ID	ORF (Annex I)	Algorithm	Amino acids
1.	38-1	AMPHI	6-14
2.	38-1	AMPHI	57-59
3.	38-1	AMPHI	67-76

4.	38-1	AMPHI	92-100
5.	38-1	AMPHI	127-137
6.	38-1	AMPHI	149-166
7.	38-1	AMPHI	210-215
8.	38-1	AMPHI	231-236
9.	38-1	AMPHI	270-272
10.	38-1	AMPHI	303-320
11.	38-1	Antigenic Index	16-34
12.	38-1	Antigenic Index	37-42
13.	38-1	Antigenic Index	46-64
14.	38-1	Antigenic Index	72-91
15.	38-1	Antigenic Index	94-112
16.	38-1	Antigenic Index	114-117
17.	38-1	Antigenic Index	124-136
18.	38-1	Antigenic Index	143-146
19.	38-1	Antigenic Index	148-160
20.	38-1	Antigenic Index	167-195
21.	38-1	Antigenic Index	201-216
22.	38-1	Antigenic Index	218-240
23.	38-1	Antigenic Index	244-252
24.	38-1	Antigenic Index	257-278
25.	38-1	Antigenic Index	282-290
26.	38-1	Antigenic Index	308-314
27.	38-1	Hydrophilicity	21-34
28.	38-1	Hydrophilicity	37-42
29.	38-1	Hydrophilicity	47-55
30.	38-1	Hydrophilicity	57-61
31.	38-1	Hydrophilicity	72-74
32.	38-1	Hydrophilicity	76-78
33.	38-1	Hydrophilicity	82-91
34.	38-1	Hydrophilicity	94-101
35.	38-1	Hydrophilicity	108-112
36.	38-1	Hydrophilicity	126-136
37.	38-1	Hydrophilicity	143-146
38.	38-1	Hydrophilicity	148-160
39.	38-1	Hydrophilicity	167-195
40.	38-1	Hydrophilicity	221-223
41.	38-1	Hydrophilicity	226-236
42.	38-1	Hydrophilicity	244-250
43.	38-1	Hydrophilicity	257-274

44.	38-1	Hydrophilicity	282-286
45.	38-1	Hydrophilicity	311-314
46.	38a	AMPHI	6-14
47.	38a	AMPHI	57-59
48.	38a	AMPHI	67-76
49.	38a	AMPHI	92-100
50.	38a	AMPHI	127-137
51.	38a	AMPHI	149-166
52.	38a	AMPHI	210-215
53.	38a	AMPHI	223-225
54.	38a	AMPHI	231-236
55.	38a	AMPHI	270-272
56.	38a	AMPHI	303-320
57.	38a	Antigenic Index	16-34
58.	38a	Antigenic Index	37-42
59.	38a	Antigenic Index	46-64
60.	38a	Antigenic Index	72-91
61.	38a	Antigenic Index	94-112
62.	38a	Antigenic Index	114-117
63.	38a	Antigenic Index	124-136
64.	38a	Antigenic Index	143-146
65.	38a	Antigenic Index	148-160
66.	38a	Antigenic Index	165-195
67.	38a	Antigenic Index	201-216
68.	38a	Antigenic Index	218-240
69.	38a	Antigenic Index	244-252
70.	38a	Antigenic Index	257-278
71.	38a	Antigenic Index	282-290
72.	38a	Antigenic Index	308-314
73.	38a	Hydrophilicity	21-34
74.	38a	Hydrophilicity	37-42
75.	38a	Hydrophilicity	47-55
76.	38a	Hydrophilicity	57-61
77.	38a	Hydrophilicity	72-74
78.	38a	Hydrophilicity	76-78
79.	38a	Hydrophilicity	82-91
80.	38a	Hydrophilicity	94-101
81.	38a	Hydrophilicity	108-112
82.	38a	Hydrophilicity	126-136
83.	38a	Hydrophilicity	143-146

84.	38a	Hydrophilicity	148-160
85.	38a	Hydrophilicity	165-195
86.	38a	Hydrophilicity	221-223
87.	38a	Hydrophilicity	226-236
88.	38a	Hydrophilicity	244-250
89.	38a	Hydrophilicity	257-273
90.	38a	Hydrophilicity	282-286
91.	38a	Hydrophilicity	311-314
92.	39-1	AMPHI	6-13
93.	39-1	AMPHI	21-24
94.	39-1	AMPHI	37-40
95.	39-1	AMPHI	60-75
96.	39-1	AMPHI	118-122
97.	39-1	AMPHI	134-139
98.	39-1	AMPHI	165-183
99.	39-1	AMPHI	192-195
100.	39-1	AMPHI	233-241
101.	39-1	AMPHI	247-267
102.	39-1	AMPHI	273-275
103.	39-1	AMPHI	299-308
104.	39-1	AMPHI	310-319
105.	39-1	AMPHI	322-330
106.	39-1	AMPHI	338-347
107.	39-1	AMPHI	358-364
108.	39-1	AMPHI	366-368
109.	39-1	AMPHI	376-378
110.	39-1	AMPHI	385-392
111.	39-1	AMPHI	413-416
112.	39-1	AMPHI	421-424
113.	39-1	AMPHI	429-438
114.	39-1	AMPHI	445-454
115.	39-1	AMPHI	456-458
116.	39-1	AMPHI	498-500
117.	39-1	AMPHI	512-519
118.	39-1	AMPHI	576-587
119.	39-1	AMPHI	589-600
120.	39-1	AMPHI	650-652
121.	39-1	AMPHI	670-674
122.	39-1	Antigenic Index	26-32
123.	39-1	Antigenic Index	35-45

124.	39-1	Antigenic Index	54-69
125.	39-1	Antigenic Index	79-84
126.	39-1	Antigenic Index	88-96
127.	39-1	Antigenic Index	105-110
128.	39-1	Antigenic Index	117-124
129.	39-1	Antigenic Index	152-154
130.	39-1	Antigenic Index	190-192
131.	39-1	Antigenic Index	222-231
132.	39-1	Antigenic Index	246-265
133.	39-1	Antigenic Index	292-295
134.	39-1	Antigenic Index	318-335
135.	39-1	Antigenic Index	353-362
136.	39-1	Antigenic Index	370-372
137.	39-1	Antigenic Index	402-404
138.	39-1	Antigenic Index	406-408
139.	39-1	Antigenic Index	419-421
140.	39-1	Antigenic Index	446-449
141.	39-1	Antigenic Index	453-460
142.	39-1	Antigenic Index	465-469
143.	39-1	Antigenic Index	476-487
144.	39-1	Antigenic Index	491-499
145.	39-1	Antigenic Index	505-514
146.	39-1	Antigenic Index	522-536
147.	39-1	Antigenic Index	557-567
148.	39-1	Antigenic Index	569-575
149.	39-1	Antigenic Index	577-580
150.	39-1	Antigenic Index	593-599
151.	39-1	Antigenic Index	603-619
152.	39-1	Antigenic Index	626-628
153.	39-1	Antigenic Index	634-637
154.	39-1	Antigenic Index	639-647
155.	39-1	Antigenic Index	655-658
156.	39-1	Antigenic Index	672-674
157.	39-1	Antigenic Index	677-686
158.	39-1	Antigenic Index	688-691
159.	39-1	Antigenic Index	693-699
160.	39-1	Antigenic Index	707-710
161.	39-1	Hydrophilicity	28-32
162.	39-1	Hydrophilicity	38-44
163.	39-1	Hydrophilicity	54-69

164.	39-1	Hydrophilicity	80-83
165.	39-1	Hydrophilicity	89-96
166.	39-1	Hydrophilicity	117-119
167.	39-1	Hydrophilicity	121-123
168.	39-1	Hydrophilicity	152-154
169.	39-1	Hydrophilicity	224-231
170.	39-1	Hydrophilicity	247-265
171.	39-1	Hydrophilicity	318-332
172.	39-1	Hydrophilicity	357-361
173.	39-1	Hydrophilicity	402-404
174.	39-1	Hydrophilicity	406-408
175.	39-1	Hydrophilicity	446-449
176.	39-1	Hydrophilicity	454-459
177.	39-1	Hydrophilicity	465-469
178.	39-1	Hydrophilicity	476-487
179.	39-1	Hydrophilicity	491-499
180.	39-1	Hydrophilicity	506-514
181.	39-1	Hydrophilicity	525-535
182.	39-1	Hydrophilicity	560-567
183.	39-1	Hydrophilicity	573-575
184.	39-1	Hydrophilicity	577-580
185.	39-1	Hydrophilicity	594-596
186.	39-1	Hydrophilicity	605-607
187.	39-1	Hydrophilicity	611-619
188.	39-1	Hydrophilicity	634-637
189.	39-1	Hydrophilicity	639-647
190.	39-1	Hydrophilicity	672-674
191.	39-1	Hydrophilicity	677-686
192.	39-1	Hydrophilicity	688-690
193.	39-1	Hydrophilicity	693-695
194.	39a	AMPHI	6-13
195.	39a	AMPHI	21-24
196.	39a	AMPHI	37-40
197.	39a	AMPHI	60-75
198.	39a	AMPHI	118-122
199.	39a	AMPHI	134-139
200.	39a	AMPHI	165-183
201.	39a	AMPHI	192-195
202.	39a	AMPHI	233-241
203.	39a	AMPHI	247-267

204.	39a	AMPHI	273-275
205.	39a	AMPHI	299-308
206.	39a	AMPHI	310-319
207.	39a	AMPHI	322-330
208.	39a	AMPHI	338-347
209.	39a	AMPHI	358-364
210.	39a	AMPHI	366-368
211.	39a	AMPHI	376-378
212.	39a	AMPHI	385-392
213.	39a	AMPHI	413-416
214.	39a	AMPHI	421-424
215.	39a	AMPHI	429-438
216.	39a	AMPHI	445-454
217.	39a	AMPHI	456-458
218.	39a	AMPHI	498-500
219.	39a	AMPHI	512-520
220.	39a	AMPHI	576-587
221.	39a	AMPHI	589-600
222.	39a	AMPHI	650-652
223.	39a	AMPHI	670-674
224.	39a	Antigenic Index	26-32
225.	39a	Antigenic Index	35-45
226.	39a	Antigenic Index	54-69
227.	39a	Antigenic Index	79-84
228.	39a	Antigenic Index	89-96
229.	39a	Antigenic Index	103-110
230.	39a	Antigenic Index	117-124
231.	39a	Antigenic Index	152-154
232.	39a	Antigenic Index	190-192
233.	39a	Antigenic Index	222-231
234.	39a	Antigenic Index	246-265
235.	39a	Antigenic Index	292-295
236.	39a	Antigenic Index	318-335
237.	39a	Antigenic Index	353-362
238.	39a	Antigenic Index	370-372
239.	39a	Antigenic Index	402-404
240.	39a	Antigenic Index	406-408
241.	39a	Antigenic Index	419-421
242.	39a	Antigenic Index	446-449
243.	39a	Antigenic Index	453-460

244.	39a	Antigenic Index	465-469
245.	39a	Antigenic Index	476-487
246.	39a	Antigenic Index	491-499
247.	39a	Antigenic Index	505-514
248.	39a	Antigenic Index	529-535
249.	39a	Antigenic Index	557-567
250.	39a	Antigenic Index	569-575
251.	39a	Antigenic Index	577-580
252.	39a	Antigenic Index	593-599
253.	39a	Antigenic Index	603-619
254.	39a	Antigenic Index	626-628
255.	39a	Antigenic Index	634-637
256.	39a	Antigenic Index	639-647
257.	39a	Antigenic Index	655-658
258.	39a	Antigenic Index	672-674
259.	39a	Antigenic Index	677-686
260.	39a	Antigenic Index	688-691
261.	39a	Antigenic Index	693-699
262.	39a	Antigenic Index	707-710
263.	39a	Hydrophilicity	28-32
264.	39a	Hydrophilicity	38-44
265.	39a	Hydrophilicity	54-69
266.	39a	Hydrophilicity	80-83
267.	39a	Hydrophilicity	89-95
268.	39a	Hydrophilicity	105-108
269.	39a	Hydrophilicity	117-119
270.	39a	Hydrophilicity	121-123
271.	39a	Hydrophilicity	152-154
272.	39a	Hydrophilicity	224-231
273.	39a	Hydrophilicity	247-265
274.	39a	Hydrophilicity	318-332
275.	39a	Hydrophilicity	357-361
276.	39a	Hydrophilicity	402-404
277.	39a	Hydrophilicity	406-408
278.	39a	Hydrophilicity	446-449
279.	39a	Hydrophilicity	454-459
280.	39a	Hydrophilicity	465-469
281.	39a	Hydrophilicity	476-487
282.	39a	Hydrophilicity	491-499
283.	39a	Hydrophilicity	506-514

284.	39a	Hydrophilicity	529-535
285.	39a	Hydrophilicity	560-567
286.	39a	Hydrophilicity	573-575
287.	39a	Hydrophilicity	577-580
288.	39a	Hydrophilicity	594-596
289.	39a	Hydrophilicity	605-607
290.	39a	Hydrophilicity	611-619
291.	39a	Hydrophilicity	634-637
292.	39a	Hydrophilicity	639-647
293.	39a	Hydrophilicity	672-674
294.	39a	Hydrophilicity	677-686
295.	39a	Hydrophilicity	688-690
296.	39a	Hydrophilicity	693-695
297.	40-1	AMPHI	6-14
298.	40-1	AMPHI	16-19
299.	40-1	AMPHI	22-27
300.	40-1	AMPHI	30-33
301.	40-1	AMPHI	41-44
302.	40-1	AMPHI	62-68
303.	40-1	AMPHI	129-139
304.	40-1	AMPHI	161-165
305.	40-1	AMPHI	181-191
306.	40-1	AMPHI	199-202
307.	40-1	AMPHI	215-220
308.	40-1	AMPHI	237-249
309.	40-1	AMPHI	298-302
310.	40-1	AMPHI	313-318
311.	40-1	AMPHI	335-342
312.	40-1	AMPHI	376-383
313.	40-1	AMPHI	399-402
314.	40-1	AMPHI	426-428
315.	40-1	AMPHI	430-433
316.	40-1	AMPHI	435-437
317.	40-1	AMPHI	479-482
318.	40-1	AMPHI	491-511
319.	40-1	AMPHI	523-525
320.	40-1	AMPHI	560-563
321.	40-1	Antigenic Index	21-32
322.	40-1	Antigenic Index	49-61
323.	40-1	Antigenic Index	64-66

324.	40-1	Antigenic Index	74-92
325.	40-1	Antigenic Index	98-123
326.	40-1	Antigenic Index	129-135
327.	40-1	Antigenic Index	138-176
328.	40-1	Antigenic Index	193-195
329.	40-1	Antigenic Index	199-219
330.	40-1	Antigenic Index	226-240
331.	40-1	Antigenic Index	242-245
332.	40-1	Antigenic Index	251-257
333.	40-1	Antigenic Index	261-276
334.	40-1	Antigenic Index	279-306
335.	40-1	Antigenic Index	308-346
336.	40-1	Antigenic Index	352-367
337.	40-1	Antigenic Index	375-378
338.	40-1	Antigenic Index	384-406
339.	40-1	Antigenic Index	408-420
340.	40-1	Antigenic Index	423-426
341.	40-1	Antigenic Index	428-438
342.	40-1	Antigenic Index	453-459
343.	40-1	Antigenic Index	462-481
344.	40-1	Antigenic Index	485-494
345.	40-1	Antigenic Index	506-518
346.	40-1	Antigenic Index	535-539
347.	40-1	Antigenic Index	544-552
348.	40-1	Antigenic Index	559-566
349.	40-1	Antigenic Index	571-582
350.	40-1	Hydrophilicity	21-32
351.	40-1	Hydrophilicity	51-61
352.	40-1	Hydrophilicity	64-66
353.	40-1	Hydrophilicity	75-92
354.	40-1	Hydrophilicity	100-122
355.	40-1	Hydrophilicity	129-135
356.	40-1	Hydrophilicity	140-145
357.	40-1	Hydrophilicity	149-152
358.	40-1	Hydrophilicity	157-161
359.	40-1	Hydrophilicity	163-175
360.	40-1	Hydrophilicity	199-201
361.	40-1	Hydrophilicity	203-219
362.	40-1	Hydrophilicity	227-240
363.	40-1	Hydrophilicity	251-257

364.	40-1	Hydrophilicity	261-276
365.	40-1	Hydrophilicity	279-306
366.	40-1	Hydrophilicity	308-318
367.	40-1	Hydrophilicity	320-328
368.	40-1	Hydrophilicity	334-341
369.	40-1	Hydrophilicity	354-356
370.	40-1	Hydrophilicity	359-366
371.	40-1	Hydrophilicity	392-398
372.	40-1	Hydrophilicity	400-405
373.	40-1	Hydrophilicity	410-420
374.	40-1	Hydrophilicity	429-438
375.	40-1	Hydrophilicity	463-467
376.	40-1	Hydrophilicity	471-480
377.	40-1	Hydrophilicity	487-493
378.	40-1	Hydrophilicity	506-518
379.	40-1	Hydrophilicity	547-552
380.	40-1	Hydrophilicity	575-579
381.	40a	AMPHI	6-10
382.	40a	AMPHI	19-27
383.	40a	AMPHI	30-33
384.	40a	AMPHI	41-44
385.	40a	AMPHI	61-72
386.	40a	AMPHI	78-81
387.	40a	AMPHI	92-94
388.	40a	AMPHI	128-130
389.	40a	AMPHI	132-134
390.	40a	AMPHI	161-165
391.	40a	AMPHI	181-193
392.	40a	AMPHI	197-199
393.	40a	AMPHI	204-211
394.	40a	AMPHI	213-218
395.	40a	AMPHI	227-229
396.	40a	AMPHI	237-249
397.	40a	AMPHI	298-302
398.	40a	AMPHI	313-318
399.	40a	AMPHI	335-342
400.	40a	AMPHI	376-383
401.	40a	AMPHI	399-402
402.	40a	AMPHI	426-428
403.	40a	AMPHI	435-437

404.	40a	AMPHI	475-483
405.	40a	AMPHI	492-512
406.	40a	AMPHI	524-526
407.	40a	AMPHI	561-564
408.	40a	Antigenic Index	21-34
409.	40a	Antigenic Index	50-64
410.	40a	Antigenic Index	75-83
411.	40a	Antigenic Index	88-97
412.	40a	Antigenic Index	105-122
413.	40a	Antigenic Index	129-134
414.	40a	Antigenic Index	140-176
415.	40a	Antigenic Index	190-207
416.	40a	Antigenic Index	211-217
417.	40a	Antigenic Index	224-240
418.	40a	Antigenic Index	242-245
419.	40a	Antigenic Index	250-255
420.	40a	Antigenic Index	260-276
421.	40a	Antigenic Index	279-306
422.	40a	Antigenic Index	308-346
423.	40a	Antigenic Index	352-367
424.	40a	Antigenic Index	375-378
425.	40a	Antigenic Index	384-406
426.	40a	Antigenic Index	408-420
427.	40a	Antigenic Index	423-438
428.	40a	Antigenic Index	453-468
429.	40a	Antigenic Index	471-481
430.	40a	Antigenic Index	487-493
431.	40a	Antigenic Index	507-519
432.	40a	Antigenic Index	536-540
433.	40a	Antigenic Index	545-553
434.	40a	Antigenic Index	560-567
435.	40a	Antigenic Index	572-583
436.	40a	Hydrophilicity	21-34
437.	40a	Hydrophilicity	50-64
438.	40a	Hydrophilicity	75-83
439.	40a	Hydrophilicity	88-96
440.	40a	Hydrophilicity	105-121
441.	40a	Hydrophilicity	129-134
442.	40a	Hydrophilicity	140-145
443.	40a	Hydrophilicity	148-155

444.	40a	Hydrophilicity	157-161
445.	40a	Hydrophilicity	163-175
446.	40a	Hydrophilicity	196-202
447.	40a	Hydrophilicity	211-217
448.	40a	Hydrophilicity	225-230
449.	40a	Hydrophilicity	232-240
450.	40a	Hydrophilicity	253-255
451.	40a	Hydrophilicity	261-276
452.	40a	Hydrophilicity	279-306
453.	40a	Hydrophilicity	308-318
454.	40a	Hydrophilicity	320-328
455.	40a	Hydrophilicity	334-341
456.	40a	Hydrophilicity	354-356
457.	40a	Hydrophilicity	359-366
458.	40a	Hydrophilicity	392-398
459.	40a	Hydrophilicity	400-405
460.	40a	Hydrophilicity	410-420
461.	40a	Hydrophilicity	428-438
462.	40a	Hydrophilicity	462-468
463.	40a	Hydrophilicity	472-481
464.	40a	Hydrophilicity	489-493
465.	40a	Hydrophilicity	507-519
466.	40a	Hydrophilicity	548-553
467.	40a	Hydrophilicity	576-580
468.	41-1	AMPHI	30-36
469.	41-1	AMPHI	93-98
470.	41-1	AMPHI	111-122
471.	41-1	AMPHI	126-129
472.	41-1	AMPHI	136-143
473.	41-1	AMPHI	145-150
474.	41-1	AMPHI	156-158
475.	41-1	AMPHI	186-195
476.	41-1	AMPHI	201-208
477.	41-1	AMPHI	213-223
478.	41-1	AMPHI	236-247
479.	41-1	AMPHI	250-255
480.	41-1	AMPHI	273-282
481.	41-1	AMPHI	303-309
482.	41-1	AMPHI	311-314
483.	41-1	AMPHI	329-338

484.	41-1	AMPHI	344-362
485.	41-1	AMPHI	372-377
486.	41-1	AMPHI	385-392
487.	41-1	AMPHI	409-412
488.	41-1	AMPHI	419-426
489.	41-1	AMPHI	458-463
490.	41-1	AMPHI	470-474
491.	41-1	AMPHI	486-489
492.	41-1	AMPHI	512-518
493.	41-1	AMPHI	527-551
494.	41-1	AMPHI	564-579
495.	41-1	AMPHI	593-597
496.	41-1	Antigenic Index	13-22
497.	41-1	Antigenic Index	30-38
498.	41-1	Antigenic Index	43-55
499.	41-1	Antigenic Index	73-75
500.	41-1	Antigenic Index	87-89
501.	41-1	Antigenic Index	105-112
502.	41-1	Antigenic Index	114-124
503.	41-1	Antigenic Index	136-141
504.	41-1	Antigenic Index	147-153
505.	41-1	Antigenic Index	163-166
506.	41-1	Antigenic Index	174-184
507.	41-1	Antigenic Index	195-207
508.	41-1	Antigenic Index	226-236
509.	41-1	Antigenic Index	244-246
510.	41-1	Antigenic Index	249-265
511.	41-1	Antigenic Index	281-287
512.	41-1	Antigenic Index	294-313
513.	41-1	Antigenic Index	317-342
514.	41-1	Antigenic Index	350-375
515.	41-1	Antigenic Index	379-386
516.	41-1	Antigenic Index	390-396
517.	41-1	Antigenic Index	413-422
518.	41-1	Antigenic Index	425-430
519.	41-1	Antigenic Index	436-440
520.	41-1	Antigenic Index	446-465
521.	41-1	Antigenic Index	468-495
522.	41-1	Antigenic Index	498-518
523.	41-1	Antigenic Index	520-522

524.	41-1	Antigenic Index	525-542
525.	41-1	Antigenic Index	547-558
526.	41-1	Antigenic Index	565-590
527.	41-1	Antigenic Index	595-602
528.	41-1	Antigenic Index	608-619
529.	41-1	Hydrophilicity	14-21
530.	41-1	Hydrophilicity	30-33
531.	41-1	Hydrophilicity	45-55
532.	41-1	Hydrophilicity	87-89
533.	41-1	Hydrophilicity	106-111
534.	41-1	Hydrophilicity	114-120
535.	41-1	Hydrophilicity	122-124
536.	41-1	Hydrophilicity	136-141
537.	41-1	Hydrophilicity	148-150
538.	41-1	Hydrophilicity	177-184
539.	41-1	Hydrophilicity	195-207
540.	41-1	Hydrophilicity	226-234
541.	41-1	Hydrophilicity	249-265
542.	41-1	Hydrophilicity	285-287
543.	41-1	Hydrophilicity	294-297
544.	41-1	Hydrophilicity	299-313
545.	41-1	Hydrophilicity	317-321
546.	41-1	Hydrophilicity	323-342
547.	41-1	Hydrophilicity	350-371
548.	41-1	Hydrophilicity	379-386
549.	41-1	Hydrophilicity	417-422
550.	41-1	Hydrophilicity	425-427
551.	41-1	Hydrophilicity	447-449
552.	41-1	Hydrophilicity	459-462
553.	41-1	Hydrophilicity	468-475
554.	41-1	Hydrophilicity	479-482
555.	41-1	Hydrophilicity	484-491
556.	41-1	Hydrophilicity	499-518
557.	41-1	Hydrophilicity	520-522
558.	41-1	Hydrophilicity	526-542
559.	41-1	Hydrophilicity	550-558
560.	41-1	Hydrophilicity	568-590
561.	41-1	Hydrophilicity	595-598
562.	41-1	Hydrophilicity	617-619
563.	41a	AMPHI	6-12

564.	41a	AMPHI	32-34
565.	41a	AMPHI	69-74
566.	41a	AMPHI	86-98
567.	41a	AMPHI	111-119
568.	41a	AMPHI	121-126
569.	41a	AMPHI	132-134
570.	41a	AMPHI	155-160
571.	41a	AMPHI	162-171
572.	41a	AMPHI	177-184
573.	41a	AMPHI	189-199
574.	41a	AMPHI	212-223
575.	41a	AMPHI	226-231
576.	41a	AMPHI	249-258
577.	41a	AMPHI	287-290
578.	41a	AMPHI	305-314
579.	41a	AMPHI	320-338
580.	41a	AMPHI	348-353
581.	41a	AMPHI	361-368
582.	41a	AMPHI	385-388
583.	41a	AMPHI	395-402
584.	41a	AMPHI	434-439
585.	41a	AMPHI	446-450
586.	41a	AMPHI	462-467
587.	41a	AMPHI	470-475
588.	41a	AMPHI	488-494
589.	41a	AMPHI	503-525
590.	41a	AMPHI	540-555
591.	41a	AMPHI	569-573
592.	41a	AMPHI	578-594
593.	41a	Antigenic Index	10-13
594.	41a	Antigenic Index	19-31
595.	41a	Antigenic Index	48-50
596.	41a	Antigenic Index	63-65
597.	41a	Antigenic Index	82-101
598.	41a	Antigenic Index	112-117
599.	41a	Antigenic Index	123-129
600.	41a	Antigenic Index	139-142
601.	41a	Antigenic Index	150-160
602.	41a	Antigenic Index	171-183
603.	41a	Antigenic Index	202-212

604.	41a	Antigenic Index	220-222
605.	41a	Antigenic Index	225-241
606.	41a	Antigenic Index	257-263
607.	41a	Antigenic Index	270-289
608.	41a	Antigenic Index	293-318
609.	41a	Antigenic Index	326-351
610.	41a	Antigenic Index	355-362
611.	41a	Antigenic Index	366-372
612.	41a	Antigenic Index	389-398
613.	41a	Antigenic Index	401-406
614.	41a	Antigenic Index	412-416
615.	41a	Antigenic Index	422-441
616.	41a	Antigenic Index	444-446
617.	41a	Antigenic Index	451-471
618.	41a	Antigenic Index	475-494
619.	41a	Antigenic Index	496-498
620.	41a	Antigenic Index	501-518
621.	41a	Antigenic Index	523-534
622.	41a	Antigenic Index	540-566
623.	41a	Antigenic Index	571-578
624.	41a	Antigenic Index	582-595
625.	41a	Hydrophilicity	21-31
626.	41a	Hydrophilicity	63-65
627.	41a	Hydrophilicity	83-96
628.	41a	Hydrophilicity	98-100
629.	41a	Hydrophilicity	112-117
630.	41a	Hydrophilicity	124-126
631.	41a	Hydrophilicity	153-160
632.	41a	Hydrophilicity	171-183
633.	41a	Hydrophilicity	202-210
634.	41a	Hydrophilicity	220-222
635.	41a	Hydrophilicity	225-241
636.	41a	Hydrophilicity	261-263
637.	41a	Hydrophilicity	270-273
638.	41a	Hydrophilicity	275-289
639.	41a	Hydrophilicity	293-297
640.	41a	Hydrophilicity	299-318
641.	41a	Hydrophilicity	326-347
642.	41a	Hydrophilicity	355-362
643.	41a	Hydrophilicity	393-398

644.	41a	Hydrophilicity	401-403
645.	41a	Hydrophilicity	423-425
646.	41a	Hydrophilicity	435-438
647.	41a	Hydrophilicity	454-458
648.	41a	Hydrophilicity	460-471
649.	41a	Hydrophilicity	475-494
650.	41a	Hydrophilicity	496-498
651.	41a	Hydrophilicity	502-518
652.	41a	Hydrophilicity	527-534
653.	41a	Hydrophilicity	544-566
654.	41a	Hydrophilicity	571-574
655.	41a	Hydrophilicity	593-595
656.	44-1	AMPHI	57-60
657.	44-1	AMPHI	76-79
658.	44-1	Antigenic Index	22-34
659.	44-1	Antigenic Index	38-46
660.	44-1	Antigenic Index	50-55
661.	44-1	Antigenic Index	64-70
662.	44-1	Antigenic Index	72-80
663.	44-1	Antigenic Index	83-89
664.	44-1	Antigenic Index	96-106
665.	44-1	Antigenic Index	110-124
666.	44-1	Hydrophilicity	22-34
667.	44-1	Hydrophilicity	40-46
668.	44-1	Hydrophilicity	64-69
669.	44-1	Hydrophilicity	73-80
670.	44-1	Hydrophilicity	84-89
671.	44-1	Hydrophilicity	97-106
672.	44-1	Hydrophilicity	120-124
673.	44a	AMPHI	57-60
674.	44a	AMPHI	76-79
675.	44a	Antigenic Index	23-34
676.	44a	Antigenic Index	38-46
677.	44a	Antigenic Index	50-55
678.	44a	Antigenic Index	64-70
679.	44a	Antigenic Index	72-80
680.	44a	Antigenic Index	83-89
681.	44a	Antigenic Index	96-106
682.	44a	Antigenic Index	110-124
683.	44a	Hydrophilicity	28-34

684.	44a	Hydrophilicity	40-46
685.	44a	Hydrophilicity	64-69
686.	44a	Hydrophilicity	73-80
687.	44a	Hydrophilicity	84-89
688.	44a	Hydrophilicity	97-106
689.	44a	Hydrophilicity	120-124
690.	49-1	AMPHI	16-21
691.	49-1	AMPHI	44-48
692.	49-1	AMPHI	56-61
693.	49-1	AMPHI	92-97
694.	49-1	AMPHI	118-127
695.	49-1	AMPHI	130-149
696.	49-1	AMPHI	156-178
697.	49-1	AMPHI	235-240
698.	49-1	AMPHI	253-264
699.	49-1	AMPHI	268-271
700.	49-1	AMPHI	278-285
701.	49-1	AMPHI	287-292
702.	49-1	AMPHI	298-300
703.	49-1	AMPHI	328-337
704.	49-1	AMPHI	343-350
705.	49-1	AMPHI	355-365
706.	49-1	AMPHI	378-389
707.	49-1	AMPHI	422-424
708.	49-1	AMPHI	442-450
709.	49-1	AMPHI	464-481
710.	49-1	AMPHI	486-496
711.	49-1	AMPHI	514-521
712.	49-1	AMPHI	548-551
713.	49-1	AMPHI	553-557
714.	49-1	AMPHI	562-568
715.	49-1	AMPHI	573-575
716.	49-1	AMPHI	588-590
717.	49-1	AMPHI	603-605
718.	49-1	AMPHI	614-618
719.	49-1	Antigenic Index	15-21
720.	49-1	Antigenic Index	26-43
721.	49-1	Antigenic Index	50-59
722.	49-1	Antigenic Index	61-75
723.	49-1	Antigenic Index	79-87

724.	49-1	Antigenic Index	98-108
725.	49-1	Antigenic Index	110-120
726.	49-1	Antigenic Index	122-139
727.	49-1	Antigenic Index	147-164
728.	49-1	Antigenic Index	171-179
729.	49-1	Antigenic Index	185-197
730.	49-1	Antigenic Index	214-216
731.	49-1	Antigenic Index	229-231
732.	49-1	Antigenic Index	248-266
733.	49-1	Antigenic Index	278-283
734.	49-1	Antigenic Index	289-295
735.	49-1	Antigenic Index	316-326
736.	49-1	Antigenic Index	337-349
737.	49-1	Antigenic Index	368-378
738.	49-1	Antigenic Index	386-388
739.	49-1	Antigenic Index	390-410
740.	49-1	Antigenic Index	412-414
741.	49-1	Antigenic Index	423-429
742.	49-1	Antigenic Index	438-454
743.	49-1	Antigenic Index	462-475
744.	49-1	Antigenic Index	482-500
745.	49-1	Antigenic Index	503-509
746.	49-1	Antigenic Index	521-528
747.	49-1	Antigenic Index	540-562
748.	49-1	Antigenic Index	572-579
749.	49-1	Antigenic Index	590-606
750.	49-1	Antigenic Index	610-612
751.	49-1	Antigenic Index	617-619
752.	49-1	Antigenic Index	626-634
753.	49-1	Antigenic Index	637-640
754.	49-1	Hydrophilicity	18-21
755.	49-1	Hydrophilicity	26-29
756.	49-1	Hydrophilicity	31-43
757.	49-1	Hydrophilicity	51-57
758.	49-1	Hydrophilicity	64-68
759.	49-1	Hydrophilicity	79-87
760.	49-1	Hydrophilicity	98-107
761.	49-1	Hydrophilicity	122-125
762.	49-1	Hydrophilicity	147-164
763.	49-1	Hydrophilicity	172-175

764.	49-1	Hydrophilicity	187-197
765.	49-1	Hydrophilicity	229-231
766.	49-1	Hydrophilicity	256-262
767.	49-1	Hydrophilicity	264-266
768.	49-1	Hydrophilicity	278-283
769.	49-1	Hydrophilicity	290-292
770.	49-1	Hydrophilicity	319-326
771.	49-1	Hydrophilicity	337-349
772.	49-1	Hydrophilicity	368-376
773.	49-1	Hydrophilicity	386-388
774.	49-1	Hydrophilicity	390-410
775.	49-1	Hydrophilicity	412-414
776.	49-1	Hydrophilicity	423-429
777.	49-1	Hydrophilicity	441-451
778.	49-1	Hydrophilicity	466-472
779.	49-1	Hydrophilicity	484-490
780.	49-1	Hydrophilicity	492-494
781.	49-1	Hydrophilicity	496-498
782.	49-1	Hydrophilicity	522-528
783.	49-1	Hydrophilicity	543-562
784.	49-1	Hydrophilicity	591-606
785.	49-1	Hydrophilicity	617-619
786.	49-1	Hydrophilicity	626-632
787.	49-1	Hydrophilicity	637-640
788.	49a	AMPHI	55-61
789.	49a	AMPHI	92-97
790.	49a	AMPHI	118-127
791.	49a	AMPHI	129-135
792.	49a	AMPHI	137-145
793.	49a	AMPHI	156-178
794.	49a	AMPHI	198-200
795.	49a	AMPHI	235-240
796.	49a	AMPHI	252-264
797.	49a	AMPHI	277-285
798.	49a	AMPHI	287-292
799.	49a	AMPHI	298-300
800.	49a	AMPHI	321-326
801.	49a	AMPHI	328-337
802.	49a	AMPHI	343-350
803.	49a	AMPHI	355-365

804.	49a	AMPHI	378-389
805.	49a	AMPHI	392-397
806.	49a	AMPHI	415-424
807.	49a	AMPHI	453-456
808.	49a	AMPHI	471-480
809.	49a	AMPHI	486-504
810.	49a	AMPHI	514-519
811.	49a	AMPHI	527-534
812.	49a	AMPHI	551-554
813.	49a	AMPHI	561-568
814.	49a	AMPHI	600-605
815.	49a	AMPHI	612-616
816.	49a	AMPHI	628-633
817.	49a	AMPHI	636-641
818.	49a	AMPHI	654-660
819.	49a	AMPHI	669-691
820.	49a	AMPHI	706-721
821.	49a	AMPHI	735-739
822.	49a	AMPHI	744-760
823.	49a	Antigenic Index	4-23
824.	49a	Antigenic Index	27-43
825.	49a	Antigenic Index	51-62
826.	49a	Antigenic Index	64-68
827.	49a	Antigenic Index	72-75
828.	49a	Antigenic Index	79-87
829.	49a	Antigenic Index	98-108
830.	49a	Antigenic Index	110-120
831.	49a	Antigenic Index	124-139
832.	49a	Antigenic Index	147-164
833.	49a	Antigenic Index	176-179
834.	49a	Antigenic Index	185-197
835.	49a	Antigenic Index	214-216
836.	49a	Antigenic Index	229-231
837.	49a	Antigenic Index	248-267
838.	49a	Antigenic Index	278-283
839.	49a	Antigenic Index	289-295
840.	49a	Antigenic Index	305-308
841.	49a	Antigenic Index	316-326
842.	49a	Antigenic Index	337-349
843.	49a	Antigenic Index	368-378

844.	49a	Antigenic Index	386-388
845.	49a	Antigenic Index	391-407
846.	49a	Antigenic Index	423-429
847.	49a	Antigenic Index	436-455
848.	49a	Antigenic Index	459-484
849.	49a	Antigenic Index	492-517
850.	49a	Antigenic Index	521-528
851.	49a	Antigenic Index	532-539
852.	49a	Antigenic Index	555-564
853.	49a	Antigenic Index	567-572
854.	49a	Antigenic Index	578-582
855.	49a	Antigenic Index	588-607
856.	49a	Antigenic Index	610-612
857.	49a	Antigenic Index	617-637
858.	49a	Antigenic Index	641-660
859.	49a	Antigenic Index	662-664
860.	49a	Antigenic Index	667-684
861.	49a	Antigenic Index	689-700
862.	49a	Antigenic Index	706-732
863.	49a	Antigenic Index	737-744
864.	49a	Antigenic Index	748-761
865.	49a	Hydrophilicity	4-23
866.	49a	Hydrophilicity	31-43
867.	49a	Hydrophilicity	51-53
868.	49a	Hydrophilicity	55-57
869.	49a	Hydrophilicity	64-68
870.	49a	Hydrophilicity	79-87
871.	49a	Hydrophilicity	98-106
872.	49a	Hydrophilicity	114-120
873.	49a	Hydrophilicity	130-139
874.	49a	Hydrophilicity	147-164
875.	49a	Hydrophilicity	187-197
876.	49a	Hydrophilicity	229-231
877.	49a	Hydrophilicity	249-262
878.	49a	Hydrophilicity	264-266
879.	49a	Hydrophilicity	278-283
880.	49a	Hydrophilicity	290-292
881.	49a	Hydrophilicity	319-326
882.	49a	Hydrophilicity	337-349
883.	49a	Hydrophilicity	368-376

884.	49a	Hydrophilicity	386-388
885.	49a	Hydrophilicity	391-407
886.	49a	Hydrophilicity	427-429
887.	49a	Hydrophilicity	436-439
888.	49a	Hydrophilicity	441-455
889.	49a	Hydrophilicity	459-463
890.	49a	Hydrophilicity	465-484
891.	49a	Hydrophilicity	492-513
892.	49a	Hydrophilicity	521-528
893.	49a	Hydrophilicity	559-564
894.	49a	Hydrophilicity	567-569
895.	49a	Hydrophilicity	589-591
896.	49a	Hydrophilicity	601-604
897.	49a	Hydrophilicity	620-624
898.	49a	Hydrophilicity	626-637
899.	49a	Hydrophilicity	641-660
900.	49a	Hydrophilicity	662-664
901.	49a	Hydrophilicity	668-684
902.	49a	Hydrophilicity	693-700
903.	49a	Hydrophilicity	710-732
904.	49a	Hydrophilicity	737-740
905.	49a	Hydrophilicity	759-761
906.	51-1	AMPHI	15-21
907.	51-1	AMPHI	40-54
908.	51-1	AMPHI	75-86
909.	51-1	AMPHI	108-110
910.	51-1	AMPHI	112-124
911.	51-1	AMPHI	141-148
912.	51-1	AMPHI	184-189
913.	51-1	AMPHI	211-216
914.	51-1	Antigenic Index	58-65
915.	51-1	Antigenic Index	123-127
916.	51-1	Antigenic Index	132-137
917.	51-1	Antigenic Index	149-153
918.	51-1	Antigenic Index	165-177
919.	51-1	Antigenic Index	198-204
920.	51-1	Antigenic Index	222-231
921.	51-1	Hydrophilicity	60-65
922.	51-1	Hydrophilicity	123-127
923.	51-1	Hydrophilicity	132-135

924.	51-1	Hydrophilicity	165-174
925.	51-1	Hydrophilicity	200-203
926.	51-1	Hydrophilicity	222-227
927.	51a	AMPHI	15-21
928.	51a	AMPHI	40-54
929.	51a	AMPHI	75-86
930.	51a	AMPHI	108-110
931.	51a	AMPHI	112-124
932.	51a	AMPHI	141-148
933.	51a	AMPHI	184-189
934.	51a	AMPHI	211-216
935.	51a	Hydrophilicity	60-65
936.	51a	Hydrophilicity	123-127
937.	51a	Hydrophilicity	132-135
938.	51a	Hydrophilicity	165-174
939.	51a	Hydrophilicity	200-203
940.	51a	Hydrophilicity	222-227
941.	52-1	AMPHI	48-50
942.	52-1	AMPHI	64-73
943.	52-1	Antigenic Index	19-26
944.	52-1	Antigenic Index	30-35
945.	52-1	Antigenic Index	42-52
946.	52-1	Antigenic Index	57-86
947.	52-1	Hydrophilicity	22-26
948.	52-1	Hydrophilicity	30-35
949.	52-1	Hydrophilicity	42-52
950.	52-1	Hydrophilicity	57-71
951.	52-1	Hydrophilicity	78-86
952.	69-1	AMPHI	25-27
953.	69-1	AMPHI	46-66
954.	69-1	Antigenic Index	32-41
955.	69-1	Antigenic Index	43-45
956.	69-1	Antigenic Index	71-78
957.	69-1	Hydrophilicity	32-38
958.	69-1	Hydrophilicity	71-78
959.	69a	AMPHI	25-27
960.	69a	AMPHI	46-66
961.	69a	Antigenic Index	32-41
962.	69a	Antigenic Index	43-46
963.	69a	Antigenic Index	71-78

964.	69a	Hydrophilicity	32-38
965.	69a	Hydrophilicity	71-78
966.	77-1	AMPHI	12-16
967.	77-1	AMPHI	23-33
968.	77-1	AMPHI	35-42
969.	77-1	AMPHI	51-57
970.	77-1	AMPHI	67-70
971.	77-1	AMPHI	73-79
972.	77-1	AMPHI	122-124
973.	77-1	AMPHI	130-134
974.	77-1	AMPHI	165-178
975.	77-1	AMPHI	191-211
976.	77-1	Antigenic Index	22-31
977.	77-1	Antigenic Index	34-44
978.	77-1	Antigenic Index	80-94
979.	77-1	Antigenic Index	101-104
980.	77-1	Antigenic Index	155-158
981.	77-1	Antigenic Index	167-181
982.	77-1	Hydrophilicity	22-28
983.	77-1	Hydrophilicity	38-44
984.	77-1	Hydrophilicity	80-92
985.	77-1	Hydrophilicity	171-178
986.	77a	AMPHI	8-15
987.	77a	AMPHI	24-30
988.	77a	AMPHI	40-43
989.	77a	AMPHI	46-52
990.	77a	AMPHI	95-97
991.	77a	AMPHI	103-107
992.	77a	AMPHI	114-125
993.	77a	AMPHI	144-151
994.	77a	AMPHI	154-156
995.	77a	AMPHI	166-184
996.	77a	Antigenic Index	7-17
997.	77a	Antigenic Index	53-67
998.	77a	Antigenic Index	74-77
999.	77a	Antigenic Index	128-131
1000.	77a	Antigenic Index	140-154
1001.	77a	Hydrophilicity	11-17
1002.	77a	Hydrophilicity	53-65
1003.	77a	Hydrophilicity	141-151

1004.	81-1	AMPHI	30-40
1005.	81-1	AMPHI	54-56
1006.	81-1	AMPHI	60-63
1007.	81-1	AMPHI	76-93
1008.	81-1	AMPHI	96-101
1009.	81-1	AMPHI	104-406
1010.	81-1	AMPHI	118-126
1011.	81-1	AMPHI	190-205
1012.	81-1	AMPHI	230-233
1013.	81-1	AMPHI	239-242
1014.	81-1	AMPHI	256-258
1015.	81-1	AMPHI	264-284
1016.	81-1	AMPHI	290-297
1017.	81-1	AMPHI	317-326
1018.	81-1	AMPHI	388-396
1019.	81-1	AMPHI	403-414
1020.	81-1	AMPHI	458-463
1021.	81-1	AMPHI	476-480
1022.	81-1	Antigenic Index	1-4
1023.	81-1	Antigenic Index	35-38
1024.	81-1	Antigenic Index	86-89
1025.	81-1	Antigenic Index	95-98
1026.	81-1	Antigenic Index	100-103
1027.	81-1	Antigenic Index	128-136
1028.	81-1	Antigenic Index	154-174
1029.	81-1	Antigenic Index	197-211
1030.	81-1	Antigenic Index	220-226
1031.	81-1	Antigenic Index	232-240
1032.	81-1	Antigenic Index	244-249
1033.	81-1	Antigenic Index	251-253
1034.	81-1	Antigenic Index	255-258
1035.	81-1	Antigenic Index	276-290
1036.	81-1	Antigenic Index	292-301
1037.	81-1	Antigenic Index	307-312
1038.	81-1	Antigenic Index	318-323
1039.	81-1	Antigenic Index	334-345
1040.	81-1	Antigenic Index	352-358
1041.	81-1	Antigenic Index	364-372
1042.	81-1	Antigenic Index	376-384
1043.	81-1	Antigenic Index	387-401

1044.	81-1	Antigenic Index	409-417
1045.	81-1	Antigenic Index	423-444
1046.	81-1	Antigenic Index	452-459
1047.	81-1	Antigenic Index	486-488
1048.	81-1	Antigenic Index	490-499
1049.	81-1	Antigenic Index	507-520
1050.	81-1	Hydrophilicity	1-4
1051.	81-1	Hydrophilicity	35-38
1052.	81-1	Hydrophilicity	95-98
1053.	81-1	Hydrophilicity	128-136
1054.	81-1	Hydrophilicity	154-164
1055.	81-1	Hydrophilicity	166-172
1056.	81-1	Hydrophilicity	202-209
1057.	81-1	Hydrophilicity	220-226
1058.	81-1	Hydrophilicity	234-238
1059.	81-1	Hydrophilicity	245-249
1060.	81-1	Hydrophilicity	251-253
1061.	81-1	Hydrophilicity	284-287
1062.	81-1	Hydrophilicity	292-299
1063.	81-1	Hydrophilicity	307-312
1064.	81-1	Hydrophilicity	321-323
1065.	81-1	Hydrophilicity	338-345
1066.	81-1	Hydrophilicity	366-368
1067.	81-1	Hydrophilicity	378-384
1068.	81-1	Hydrophilicity	387-401
1069.	81-1	Hydrophilicity	409-415
1070.	81-1	Hydrophilicity	453-459
1071.	81-1	Hydrophilicity	493-499
1072.	81-1	Hydrophilicity	507-509
1073.	81-1	Hydrophilicity	512-518
1074.	82a	AMPHI	36-40
1075.	82a	AMPHI	95-111
1076.	82a	AMPHI	117-132
1077.	82a	AMPHI	135-137
1078.	82a	AMPHI	160-174
1079.	82a	AMPHI	183-187
1080.	82a	Antigenic Index	2-8
1081.	82a	Antigenic Index	56-60
1082.	82a	Antigenic Index	90-97
1083.	82a	Antigenic Index	104-111

1084.	82a	Antigenic Index	114-137
1085.	82a	Antigenic Index	141-151
1086.	82a	Antigenic Index	170-175
1087.	82a	Antigenic Index	180-188
1088.	82a	Antigenic Index	194-201
1089.	82a	Antigenic Index	206-209
1090.	82a	Antigenic Index	216-218
1091.	82a	Hydrophilicity	2-8
1092.	82a	Hydrophilicity	56-60
1093.	82a	Hydrophilicity	90-97
1094.	82a	Hydrophilicity	105-108
1095.	82a	Hydrophilicity	120-128
1096.	82a	Hydrophilicity	130-134
1097.	82a	Hydrophilicity	141-151
1098.	82a	Hydrophilicity	170-175
1099.	82a	Hydrophilicity	186-188
1100.	82a	Hydrophilicity	195-201
1101.	82a	Hydrophilicity	206-209
1102.	112-1	AMPHI	6-8
1103.	112-1	AMPHI	12-34
1104.	112-1	AMPHI	45-53
1105.	112-1	AMPHI	63-65
1106.	112-1	AMPHI	70-82
1107.	112-1	AMPHI	84-86
1108.	112-1	AMPHI	107-109
1109.	112-1	AMPHI	116-123
1110.	112-1	AMPHI	183-186
1111.	112-1	AMPHI	244-246
1112.	112-1	AMPHI	248-258
1113.	112-1	AMPHI	280-282
1114.	112-1	AMPHI	302-313
1115.	112-1	Antigenic Index	35-44
1116.	112-1	Antigenic Index	57-61
1117.	112-1	Antigenic Index	81-84
1118.	112-1	Antigenic Index	91-98
1119.	112-1	Antigenic Index	125-133
1120.	112-1	Antigenic Index	140-147
1121.	112-1	Antigenic Index	149-159
1122.	112-1	Antigenic Index	161-165
1123.	112-1	Antigenic Index	174-190

1124.	112-1	Antigenic Index	192-200
1125.	112-1	Antigenic Index	202-216
1126.	112-1	Antigenic Index	218-224
1127.	112-1	Antigenic Index	228-232
1128.	112-1	Antigenic Index	239-244
1129.	112-1	Antigenic Index	255-263
1130.	112-1	Antigenic Index	290-300
1131.	112-1	Hydrophilicity	38-40
1132.	112-1	Hydrophilicity	57-61
1133.	112-1	Hydrophilicity	92-98
1134.	112-1	Hydrophilicity	125-133
1135.	112-1	Hydrophilicity	141-143
1136.	112-1	Hydrophilicity	150-159
1137.	112-1	Hydrophilicity	161-164
1138.	112-1	Hydrophilicity	175-190
1139.	112-1	Hydrophilicity	203-216
1140.	112-1	Hydrophilicity	218-224
1141.	112-1	Hydrophilicity	228-232
1142.	112-1	Hydrophilicity	239-244
1143.	112-1	Hydrophilicity	259-261
1144.	112-1	Hydrophilicity	293-297
1145.	112a	AMPHI	6-8
1146.	112a	AMPHI	12-34
1147.	112a	AMPHI	47-54
1148.	112a	AMPHI	63-65
1149.	112a	AMPHI	69-72
1150.	112a	AMPHI	84-86
1151.	112a	AMPHI	89-91
1152.	112a	AMPHI	107-109
1153.	112a	AMPHI	116-123
1154.	112a	AMPHI	183-186
1155.	112a	AMPHI	244-246
1156.	112a	AMPHI	248-258
1157.	112a	AMPHI	280-282
1158.	112a	AMPHI	302-310
1159.	112a	AMPHI	321-336
1160.	112a	Antigenic Index	35-44
1161.	112a	Antigenic Index	57-61
1162.	112a	Antigenic Index	81-84
1163.	112a	Antigenic Index	91-98

1164.	112a	Antigenic Index	125-133
1165.	112a	Antigenic Index	140-147
1166.	112a	Antigenic Index	150-158
1167.	112a	Antigenic Index	161-164
1168.	112a	Antigenic Index	174-190
1169.	112a	Antigenic Index	194-200
1170.	112a	Antigenic Index	202-216
1171.	112a	Antigenic Index	218-220
1172.	112a	Antigenic Index	222-224
1173.	112a	Antigenic Index	228-232
1174.	112a	Antigenic Index	239-244
1175.	112a	Antigenic Index	256-263
1176.	112a	Antigenic Index	290-301
1177.	112a	Antigenic Index	351-356
1178.	112a	Hydrophilicity	38-40
1179.	112a	Hydrophilicity	57-61
1180.	112a	Hydrophilicity	93-98
1181.	112a	Hydrophilicity	125-133
1182.	112a	Hydrophilicity	141-143
1183.	112a	Hydrophilicity	150-155
1184.	112a	Hydrophilicity	161-164
1185.	112a	Hydrophilicity	175-190
1186.	112a	Hydrophilicity	203-216
1187.	112a	Hydrophilicity	218-220
1188.	112a	Hydrophilicity	222-224
1189.	112a	Hydrophilicity	228-232
1190.	112a	Hydrophilicity	239-244
1191.	112a	Hydrophilicity	259-261
1192.	112a	Hydrophilicity	293-297
1193.	112a	Hydrophilicity	351-356
1194.	114-1	AMPHI	45-54
1195.	114-1	AMPHI	154-160
1196.	114-1	AMPHI	182-190
1197.	114-1	AMPHI	224-226
1198.	114-1	AMPHI	229-233
1199.	114-1	AMPHI	285-287
1200.	114-1	AMPHI	303-310
1201.	114-1	AMPHI	321-332
1202.	114-1	AMPHI	392-398
1203.	114-1	AMPHI	413-416

1204.	114-1	AMPHI	450-452
1205.	114-1	AMPHI	477-487
1206.	114-1	AMPHI	506-509
1207.	114-1	AMPHI	525-529
1208.	114-1	AMPHI	565-567
1209.	114-1	AMPHI	614-621
1210.	114-1	AMPHI	631-635
1211.	114-1	AMPHI	770-774
1212.	114-1	AMPHI	810-813
1213.	114-1	AMPHI	847-849
1214.	114-1	AMPHI	851-853
1215.	114-1	AMPHI	875-879
1216.	114-1	AMPHI	951-956
1217.	114-1	AMPHI	975-980
1218.	114-1	AMPHI	1034-1036
1219.	114-1	AMPHI	1048-1051
1220.	114-1	AMPHI	1073-1081
1221.	114-1	AMPHI	1086-1090
1222.	114-1	AMPHI	1095-1102
1223.	114-1	AMPHI	1111-1115
1224.	114-1	AMPHI	1163-1167
1225.	114-1	AMPHI	1242-1245
1226.	114-1	AMPHI	1275-1281
1227.	114-1	AMPHI	1312-1317
1228.	114-1	AMPHI	1338-1347
1229.	114-1	AMPHI	1349-1355
1230.	114-1	AMPHI	1357-1360
1231.	114-1	AMPHI	1362-1365
1232.	114-1	AMPHI	1376-1398
1233.	114-1	AMPHI	1418-1421
1234.	114-1	AMPHI	1425-1429
1235.	114-1	AMPHI	1468-1473
1236.	114-1	AMPHI	1476-1485
1237.	114-1	AMPHI	1495-1515
1238.	114-1	AMPHI	1518-1526
1239.	114-1	AMPHI	1546-1555
1240.	114-1	AMPHI	1557-1559
1241.	114-1	AMPHI	1580-1583
1242.	114-1	AMPHI	1585-1597
1243.	114-1	AMPHI	1604-1606

1244.	114-1	AMPHI	1613-1624
1245.	114-1	AMPHI	1626-1630
1246.	114-1	AMPHI	1638-1644
1247.	114-1	AMPHI	1655-1660
1248.	114-1	AMPHI	1662-1664
1249.	114-1	AMPHI	1672-1674
1250.	114-1	AMPHI	1677-1679
1251.	114-1	AMPHI	1691-1694
1252.	114-1	AMPHI	1713-1716
1253.	114-1	AMPHI	1719-1729
1254.	114-1	AMPHI	1735-1738
1255.	114-1	AMPHI	1753-1757
1256.	114-1	AMPHI	1772-1778
1257.	114-1	AMPHI	1790-1792
1258.	114-1	AMPHI	1817-1826
1259.	114-1	AMPHI	1828-1832
1260.	114-1	AMPHI	1840-1851
1261.	114-1	AMPHI	1854-1856
1262.	114-1	AMPHI	1871-1881
1263.	114-1	AMPHI	1883-1896
1264.	114-1	AMPHI	1922-1927
1265.	114-1	AMPHI	1934-1946
1266.	114-1	AMPHI	1950-1955
1267.	114-1	AMPHI	1957-1964
1268.	114-1	Antigenic Index	1-6
1269.	114-1	Antigenic Index	10-16
1270.	114-1	Antigenic Index	23-37
1271.	114-1	Antigenic Index	41-55
1272.	114-1	Antigenic Index	75-85
1273.	114-1	Antigenic Index	91-97
1274.	114-1	Antigenic Index	102-140
1275.	114-1	Antigenic Index	147-156
1276.	114-1	Antigenic Index	161-168
1277.	114-1	Antigenic Index	172-174
1278.	114-1	Antigenic Index	181-189
1279.	114-1	Antigenic Index	196-203
1280.	114-1	Antigenic Index	208-213
1281.	114-1	Antigenic Index	220-229
1282.	114-1	Antigenic Index	242-248
1283.	114-1	Antigenic Index	251-266

1284.	114-1	Antigenic Index	268-276
1285.	114-1	Antigenic Index	295-307
1286.	114-1	Antigenic Index	309-312
1287.	114-1	Antigenic Index	318-340
1288.	114-1	Antigenic Index	345-351
1289.	114-1	Antigenic Index	357-366
1290.	114-1	Antigenic Index	371-381
1291.	114-1	Antigenic Index	385-392
1292.	114-1	Antigenic Index	404-417
1293.	114-1	Antigenic Index	419-432
1294.	114-1	Antigenic Index	440-456
1295.	114-1	Antigenic Index	464-468
1296.	114-1	Antigenic Index	473-480
1297.	114-1	Antigenic Index	482-488
1298.	114-1	Antigenic Index	496-511
1299.	114-1	Antigenic Index	515-530
1300.	114-1	Antigenic Index	535-549
1301.	114-1	Antigenic Index	555-560
1302.	114-1	Antigenic Index	564-582
1303.	114-1	Antigenic Index	588-596
1304.	114-1	Antigenic Index	602-615
1305.	114-1	Antigenic Index	617-620
1306.	114-1	Antigenic Index	622-624
1307.	114-1	Antigenic Index	628-632
1308.	114-1	Antigenic Index	637-640
1309.	114-1	Antigenic Index	647-654
1310.	114-1	Antigenic Index	660-666
1311.	114-1	Antigenic Index	668-688
1312.	114-1	Antigenic Index	696-725
1313.	114-1	Antigenic Index	730-733
1314.	114-1	Antigenic Index	738-755
1315.	114-1	Antigenic Index	760-766
1316.	114-1	Antigenic Index	779-783
1317.	114-1	Antigenic Index	786-799
1318.	114-1	Antigenic Index	807-809
1319.	114-1	Antigenic Index	811-819
1320.	114-1	Antigenic Index	831-839
1321.	114-1	Antigenic Index	845-857
1322.	114-1	Antigenic Index	860-862
1323.	114-1	Antigenic Index	864-868

1324.	114-1	Antigenic Index	872-879
1325.	114-1	Antigenic Index	883-891
1326.	114-1	Antigenic Index	893-903
1327.	114-1	Antigenic Index	908-916
1328.	114-1	Antigenic Index	919-936
1329.	114-1	Antigenic Index	941-947
1330.	114-1	Antigenic Index	950-956
1331.	114-1	Antigenic Index	959-976
1332.	114-1	Antigenic Index	979-991
1333.	114-1	Antigenic Index	993-1000
1334.	114-1	Antigenic Index	1007-1022
1335.	114-1	Antigenic Index	1041-1053
1336.	114-1	Antigenic Index	1062-1068
1337.	114-1	Antigenic Index	1075-1108
1338.	114-1	Antigenic Index	1115-1121
1339.	114-1	Antigenic Index	1126-1145
1340.	114-1	Antigenic Index	1148-1152
1341.	114-1	Antigenic Index	1156-1178
1342.	114-1	Antigenic Index	1195-1206
1343.	114-1	Antigenic Index	1208-1212
1344.	114-1	Antigenic Index	1217-1243
1345.	114-1	Antigenic Index	1246-1263
1346.	114-1	Antigenic Index	1271-1282
1347.	114-1	Antigenic Index	1284-1288
1348.	114-1	Antigenic Index	1292-1295
1349.	114-1	Antigenic Index	1299-1307
1350.	114-1	Antigenic Index	1318-1328
1351.	114-1	Antigenic Index	1330-1340
1352.	114-1	Antigenic Index	1344-1359
1353.	114-1	Antigenic Index	1367-1384
1354.	114-1	Antigenic Index	1395-1399
1355.	114-1	Antigenic Index	1405-1417
1356.	114-1	Antigenic Index	1445-1449
1357.	114-1	Antigenic Index	1491-1510
1358.	114-1	Antigenic Index	1526-1529
1359.	114-1	Antigenic Index	1532-1548
1360.	114-1	Antigenic Index	1552-1556
1361.	114-1	Antigenic Index	1560-1562
1362.	114-1	Antigenic Index	1573-1583
1363.	114-1	Antigenic Index	1594-1611

1364.	114-1	Antigenic Index	1627-1635
1365.	114-1	Antigenic Index	1643-1645
1366.	114-1	Antigenic Index	1647-1665
1367.	114-1	Antigenic Index	1680-1686
1368.	114-1	Antigenic Index	1700-1722
1369.	114-1	Antigenic Index	1724-1726
1370.	114-1	Antigenic Index	1739-1746
1371.	114-1	Antigenic Index	1752-1757
1372.	114-1	Antigenic Index	1780-1783
1373.	114-1	Antigenic Index	1791-1795
1374.	114-1	Antigenic Index	1804-1808
1375.	114-1	Antigenic Index	1829-1835
1376.	114-1	Antigenic Index	1841-1859
1377.	114-1	Antigenic Index	1867-1886
1378.	114-1	Antigenic Index	1897-1903
1379.	114-1	Antigenic Index	1908-1912
1380.	114-1	Antigenic Index	1917-1922
1381.	114-1	Antigenic Index	1926-1934
1382.	114-1	Antigenic Index	1938-1945
1383.	114-1	Antigenic Index	1947-1957
1384.	114-1	Antigenic Index	1961-1968
1385.	114-1	Antigenic Index	1974-1978
1386.	114-1	Hydrophilicity	4-6
1387.	114-1	Hydrophilicity	12-15
1388.	114-1	Hydrophilicity	23-34
1389.	114-1	Hydrophilicity	43-55
1390.	114-1	Hydrophilicity	76-85
1391.	114-1	Hydrophilicity	104-110
1392.	114-1	Hydrophilicity	118-123
1393.	114-1	Hydrophilicity	127-132
1394.	114-1	Hydrophilicity	147-154
1395.	114-1	Hydrophilicity	163-167
1396.	114-1	Hydrophilicity	185-187
1397.	114-1	Hydrophilicity	197-203
1398.	114-1	Hydrophilicity	208-211
1399.	114-1	Hydrophilicity	221-227
1400.	114-1	Hydrophilicity	243-245
1401.	114-1	Hydrophilicity	253-261
1402.	114-1	Hydrophilicity	263-266
1403.	114-1	Hydrophilicity	270-272

1404.	114-1	Hydrophilicity	295-301
1405.	114-1	Hydrophilicity	309-312
1406.	114-1	Hydrophilicity	320-328
1407.	114-1	Hydrophilicity	332-337
1408.	114-1	Hydrophilicity	345-351
1409.	114-1	Hydrophilicity	360-366
1410.	114-1	Hydrophilicity	371-378
1411.	114-1	Hydrophilicity	387-392
1412.	114-1	Hydrophilicity	404-415
1413.	114-1	Hydrophilicity	419-432
1414.	114-1	Hydrophilicity	441-450
1415.	114-1	Hydrophilicity	452-456
1416.	114-1	Hydrophilicity	473-480
1417.	114-1	Hydrophilicity	482-485
1418.	114-1	Hydrophilicity	496-500
1419.	114-1	Hydrophilicity	504-509
1420.	114-1	Hydrophilicity	515-520
1421.	114-1	Hydrophilicity	536-549
1422.	114-1	Hydrophilicity	555-560
1423.	114-1	Hydrophilicity	565-568
1424.	114-1	Hydrophilicity	570-579
1425.	114-1	Hydrophilicity	589-594
1426.	114-1	Hydrophilicity	602-604
1427.	114-1	Hydrophilicity	609-615
1428.	114-1	Hydrophilicity	617-620
1429.	114-1	Hydrophilicity	660-666
1430.	114-1	Hydrophilicity	668-680
1431.	114-1	Hydrophilicity	684-686
1432.	114-1	Hydrophilicity	699-708
1433.	114-1	Hydrophilicity	715-725
1434.	114-1	Hydrophilicity	730-733
1435.	114-1	Hydrophilicity	738-744
1436.	114-1	Hydrophilicity	746-754
1437.	114-1	Hydrophilicity	760-766
1438.	114-1	Hydrophilicity	789-793
1439.	114-1	Hydrophilicity	816-818
1440.	114-1	Hydrophilicity	831-836
1441.	114-1	Hydrophilicity	845-857
1442.	114-1	Hydrophilicity	860-862
1443.	114-1	Hydrophilicity	864-866

1444.	114-1	Hydrophilicity	873-879
1445.	114-1	Hydrophilicity	883-885
1446.	114-1	Hydrophilicity	887-889
1447.	114-1	Hydrophilicity	896-899
1448.	114-1	Hydrophilicity	908-916
1449.	114-1	Hydrophilicity	919-932
1450.	114-1	Hydrophilicity	941-947
1451.	114-1	Hydrophilicity	962-975
1452.	114-1	Hydrophilicity	979-989
1453.	114-1	Hydrophilicity	993-1000
1454.	114-1	Hydrophilicity	1007-1022
1455.	114-1	Hydrophilicity	1041-1043
1456.	114-1	Hydrophilicity	1045-1053
1457.	114-1	Hydrophilicity	1062-1068
1458.	114-1	Hydrophilicity	1075-1078
1459.	114-1	Hydrophilicity	1080-1087
1460.	114-1	Hydrophilicity	1089-1104
1461.	114-1	Hydrophilicity	1115-1121
1462.	114-1	Hydrophilicity	1126-1141
1463.	114-1	Hydrophilicity	1143-1145
1464.	114-1	Hydrophilicity	1148-1151
1465.	114-1	Hydrophilicity	1157-1178
1466.	114-1	Hydrophilicity	1197-1203
1467.	114-1	Hydrophilicity	1217-1243
1468.	114-1	Hydrophilicity	1246-1263
1469.	114-1	Hydrophilicity	1271-1273
1470.	114-1	Hydrophilicity	1275-1277
1471.	114-1	Hydrophilicity	1284-1288
1472.	114-1	Hydrophilicity	1299-1307
1473.	114-1	Hydrophilicity	1318-1326
1474.	114-1	Hydrophilicity	1334-1340
1475.	114-1	Hydrophilicity	1350-1355
1476.	114-1	Hydrophilicity	1357-1359
1477.	114-1	Hydrophilicity	1367-1384
1478.	114-1	Hydrophilicity	1407-1417
1479.	114-1	Hydrophilicity	1491-1510
1480.	114-1	Hydrophilicity	1534-1540
1481.	114-1	Hydrophilicity	1576-1583
1482.	114-1	Hydrophilicity	1595-1607
1483.	114-1	Hydrophilicity	1629-1635

1484.	114-1	Hydrophilicity	1643-1645
1485.	114-1	Hydrophilicity	1649-1665
1486.	114-1	Hydrophilicity	1682-1686
1487.	114-1	Hydrophilicity	1704-1722
1488.	114-1	Hydrophilicity	1724-1726
1489.	114-1	Hydrophilicity	1740-1746
1490.	114-1	Hydrophilicity	1804-1806
1491.	114-1	Hydrophilicity	1829-1835
1492.	114-1	Hydrophilicity	1842-1855
1493.	114-1	Hydrophilicity	1876-1879
1494.	114-1	Hydrophilicity	1898-1900
1495.	114-1	Hydrophilicity	1910-1912
1496.	114-1	Hydrophilicity	1920-1922
1497.	114-1	Hydrophilicity	1928-1930
1498.	114-1	Hydrophilicity	1938-1940
1499.	114-1	Hydrophilicity	1948-1954
1500.	114-1	Hydrophilicity	1962-1967
1501.	114a	AMPHI	45-54
1502.	114a	AMPHI	154-160
1503.	114a	AMPHI	182-190
1504.	114a	AMPHI	224-226
1505.	114a	AMPHI	229-233
1506.	114a	AMPHI	285-287
1507.	114a	AMPHI	303-310
1508.	114a	AMPHI	321-332
1509.	114a	AMPHI	348-350
1510.	114a	AMPHI	392-398
1511.	114a	AMPHI	414-416
1512.	114a	AMPHI	478-486
1513.	114a	AMPHI	506-509
1514.	114a	AMPHI	525-529
1515.	114a	AMPHI	565-567
1516.	114a	AMPHI	614-621
1517.	114a	AMPHI	631-635
1518.	114a	AMPHI	770-774
1519.	114a	AMPHI	811-813
1520.	114a	AMPHI	847-849
1521.	114a	AMPHI	851-853
1522.	114a	AMPHI	875-879
1523.	114a	AMPHI	951-959

1524.	114a	AMPHI	975-981
1525.	114a	AMPHI	1034-1036
1526.	114a	AMPHI	1048-1051
1527.	114a	AMPHI	1073-1081
1528.	114a	AMPHI	1086-1090
1529.	114a	AMPHI	1095-1102
1530.	114a	AMPHI	1111-1115
1531.	114a	AMPHI	1163-1166
1532.	114a	AMPHI	1275-1281
1533.	114a	AMPHI	1312-1317
1534.	114a	AMPHI	1338-1347
1535.	114a	AMPHI	1349-1355
1536.	114a	AMPHI	1357-1365
1537.	114a	AMPHI	1376-1398
1538.	114a	AMPHI	1418-1420
1539.	114a	AMPHI	1455-1460
1540.	114a	AMPHI	1472-1484
1541.	114a	AMPHI	1497-1505
1542.	114a	AMPHI	1507-1512
1543.	114a	Antigenic Index	1-6
1544.	114a	Antigenic Index	10-16
1545.	114a	Antigenic Index	23-37
1546.	114a	Antigenic Index	41-55
1547.	114a	Antigenic Index	75-85
1548.	114a	Antigenic Index	91-97
1549.	114a	Antigenic Index	102-137
1550.	114a	Antigenic Index	147-156
1551.	114a	Antigenic Index	161-168
1552.	114a	Antigenic Index	172-174
1553.	114a	Antigenic Index	181-189
1554.	114a	Antigenic Index	196-203
1555.	114a	Antigenic Index	208-213
1556.	114a	Antigenic Index	220-229
1557.	114a	Antigenic Index	242-248
1558.	114a	Antigenic Index	251-266
1559.	114a	Antigenic Index	268-276
1560.	114a	Antigenic Index	295-307
1561.	114a	Antigenic Index	309-312
1562.	114a	Antigenic Index	318-340
1563.	114a	Antigenic Index	345-352

1564.	114a	Antigenic Index	357-366
1565.	114a	Antigenic Index	371-381
1566.	114a	Antigenic Index	385-392
1567.	114a	Antigenic Index	404-427
1568.	114a	Antigenic Index	429-434
1569.	114a	Antigenic Index	440-456
1570.	114a	Antigenic Index	465-468
1571.	114a	Antigenic Index	473-494
1572.	114a	Antigenic Index	496-510
1573.	114a	Antigenic Index	515-530
1574.	114a	Antigenic Index	535-549
1575.	114a	Antigenic Index	555-560
1576.	114a	Antigenic Index	564-578
1577.	114a	Antigenic Index	588-596
1578.	114a	Antigenic Index	602-615
1579.	114a	Antigenic Index	617-620
1580.	114a	Antigenic Index	622-624
1581.	114a	Antigenic Index	628-632
1582.	114a	Antigenic Index	637-640
1583.	114a	Antigenic Index	647-654
1584.	114a	Antigenic Index	660-666
1585.	114a	Antigenic Index	668-688
1586.	114a	Antigenic Index	697-725
1587.	114a	Antigenic Index	730-733
1588.	114a	Antigenic Index	738-755
1589.	114a	Antigenic Index	760-766
1590.	114a	Antigenic Index	779-783
1591.	114a	Antigenic Index	786-799
1592.	114a	Antigenic Index	806-809
1593.	114a	Antigenic Index	811-819
1594.	114a	Antigenic Index	831-839
1595.	114a	Antigenic Index	845-857
1596.	114a	Antigenic Index	860-862
1597.	114a	Antigenic Index	864-868
1598.	114a	Antigenic Index	872-879
1599.	114a	Antigenic Index	883-891
1600.	114a	Antigenic Index	893-902
1601.	114a	Antigenic Index	908-916
1602.	114a	Antigenic Index	923-936
1603.	114a	Antigenic Index	941-947

1604.	114a	Antigenic Index	950-956
1605.	114a	Antigenic Index	959-976
1606.	114a	Antigenic Index	979-989
1607.	114a	Antigenic Index	993-1000
1608.	114a	Antigenic Index	1007-1022
1609.	114a	Antigenic Index	1041-1053
1610.	114a	Antigenic Index	1062-1068
1611.	114a	Antigenic Index	1075-1108
1612.	114a	Antigenic Index	1115-1121
1613.	114a	Antigenic Index	1126-1145
1614.	114a	Antigenic Index	1148-1152
1615.	114a	Antigenic Index	1157-1176
1616.	114a	Antigenic Index	1195-1206
1617.	114a	Antigenic Index	1208-1212
1618.	114a	Antigenic Index	1224-1243
1619.	114a	Antigenic Index	1247-1263
1620.	114a	Antigenic Index	1271-1282
1621.	114a	Antigenic Index	1284-1288
1622.	114a	Antigenic Index	1292-1295
1623.	114a	Antigenic Index	1299-1307
1624.	114a	Antigenic Index	1318-1328
1625.	114a	Antigenic Index	1330-1340
1626.	114a	Antigenic Index	1344-1359
1627.	114a	Antigenic Index	1367-1384
1628.	114a	Antigenic Index	1396-1399
1629.	114a	Antigenic Index	1405-1417
1630.	114a	Antigenic Index	1434-1436
1631.	114a	Antigenic Index	1449-1451
1632.	114a	Antigenic Index	1468-1487
1633.	114a	Antigenic Index	1498-1503
1634.	114a	Antigenic Index	1509-1515
1635.	114a	Antigenic Index	1525-1532
1636.	114a	Hydrophilicity	4-6
1637.	114a	Hydrophilicity	12-15
1638.	114a	Hydrophilicity	23-34
1639.	114a	Hydrophilicity	43-55
1640.	114a	Hydrophilicity	75-85
1641.	114a	Hydrophilicity	104-110
1642.	114a	Hydrophilicity	118-123
1643.	114a	Hydrophilicity	127-132

1644.	114a	Hydrophilicity	147-154
1645.	114a	Hydrophilicity	163-167
1646.	114a	Hydrophilicity	185-187
1647.	114a	Hydrophilicity	197-203
1648.	114a	Hydrophilicity	208-211
1649.	114a	Hydrophilicity	221-227
1650.	114a	Hydrophilicity	243-245
1651.	114a	Hydrophilicity	253-261
1652.	114a	Hydrophilicity	263-266
1653.	114a	Hydrophilicity	270-272
1654.	114a	Hydrophilicity	295-301
1655.	114a	Hydrophilicity	309-312
1656.	114a	Hydrophilicity	320-328
1657.	114a	Hydrophilicity	332-337
1658.	114a	Hydrophilicity	345-351
1659.	114a	Hydrophilicity	360-366
1660.	114a	Hydrophilicity	371-378
1661.	114a	Hydrophilicity	387-392
1662.	114a	Hydrophilicity	404-417
1663.	114a	Hydrophilicity	421-423
1664.	114a	Hydrophilicity	425-427
1665.	114a	Hydrophilicity	442-456
1666.	114a	Hydrophilicity	473-488
1667.	114a	Hydrophilicity	499-509
1668.	114a	Hydrophilicity	515-520
1669.	114a	Hydrophilicity	536-549
1670.	114a	Hydrophilicity	555-560
1671.	114a	Hydrophilicity	565-568
1672.	114a	Hydrophilicity	570-578
1673.	114a	Hydrophilicity	589-594
1674.	114a	Hydrophilicity	602-604
1675.	114a	Hydrophilicity	609-615
1676.	114a	Hydrophilicity	617-620
1677.	114a	Hydrophilicity	660-665
1678.	114a	Hydrophilicity	668-680
1679.	114a	Hydrophilicity	684-686
1680.	114a	Hydrophilicity	699-708
1681.	114a	Hydrophilicity	715-725
1682.	114a	Hydrophilicity	730-733
1683.	114a	Hydrophilicity	738-744

1684.	114a	Hydrophilicity	746-754
1685.	114a	Hydrophilicity	760-766
1686.	114a	Hydrophilicity	789-793
1687.	114a	Hydrophilicity	816-818
1688.	114a	Hydrophilicity	831-836
1689.	114a	Hydrophilicity	845-857
1690.	114a	Hydrophilicity	860-862
1691.	114a	Hydrophilicity	864-866
1692.	114a	Hydrophilicity	873-879
1693.	114a	Hydrophilicity	883-885
1694.	114a	Hydrophilicity	887-889
1695.	114a	Hydrophilicity	896-899
1696.	114a	Hydrophilicity	908-916
1697.	114a	Hydrophilicity	923-932
1698.	114a	Hydrophilicity	941-947
1699.	114a	Hydrophilicity	961-975
1700.	114a	Hydrophilicity	979-989
1701.	114a	Hydrophilicity	993-1000
1702.	114a	Hydrophilicity	1007-1022
1703.	114a	Hydrophilicity	1041-1043
1704.	114a	Hydrophilicity	1045-1053
1705.	114a	Hydrophilicity	1062-1068
1706.	114a	Hydrophilicity	1075-1078
1707.	114a	Hydrophilicity	1080-1087
1708.	114a	Hydrophilicity	1089-1104
1709.	114a	Hydrophilicity	1115-1121
1710.	114a	Hydrophilicity	1126-1141
1711.	114a	Hydrophilicity	1143-1145
1712.	114a	Hydrophilicity	1148-1151
1713.	114a	Hydrophilicity	1158-1171
1714.	114a	Hydrophilicity	1197-1203
1715.	114a	Hydrophilicity	1224-1243
1716.	114a	Hydrophilicity	1251-1263
1717.	114a	Hydrophilicity	1271-1273
1718.	114a	Hydrophilicity	1275-1277
1719.	114a	Hydrophilicity	1284-1288
1720.	114a	Hydrophilicity	1299-1307
1721.	114a	Hydrophilicity	1318-1326
1722.	114a	Hydrophilicity	1334-1340
1723.	114a	Hydrophilicity	1350-1359

1724.	114a	Hydrophilicity	1367-1384
1725.	114a	Hydrophilicity	1407-1417
1726.	114a	Hydrophilicity	1449-1451
1727.	114a	Hydrophilicity	1469-1482
1728.	114a	Hydrophilicity	1484-1486
1729.	114a	Hydrophilicity	1498-1503
1730.	114a	Hydrophilicity	1510-1512
1731.	114a	Hydrophilicity	1527-1532
1732.	124-1	AMPHI	37-43
1733.	124-1	AMPHI	94-96
1734.	124-1	AMPHI	113-115
1735.	124-1	Antigenic Index	20-26
1736.	124-1	Antigenic Index	38-43
1737.	124-1	Antigenic Index	52-55
1738.	124-1	Antigenic Index	62-70
1739.	124-1	Antigenic Index	88-97
1740.	124-1	Antigenic Index	104-114
1741.	124-1	Antigenic Index	123-135
1742.	124-1	Antigenic Index	146-155
1743.	124-1	Hydrophilicity	20-26
1744.	124-1	Hydrophilicity	41-43
1745.	124-1	Hydrophilicity	52-55
1746.	124-1	Hydrophilicity	63-69
1747.	124-1	Hydrophilicity	91-94
1748.	124-1	Hydrophilicity	104-114
1749.	124-1	Hydrophilicity	123-135
1750.	124-1	Hydrophilicity	146-155
1751.	124a	AMPHI	19-21
1752.	124a	AMPHI	23-29
1753.	124a	AMPHI	37-43
1754.	124a	AMPHI	94-96
1755.	124a	Antigenic Index	38-43
1756.	124a	Antigenic Index	52-55
1757.	124a	Antigenic Index	62-70
1758.	124a	Antigenic Index	77-80
1759.	124a	Antigenic Index	90-96
1760.	124a	Antigenic Index	105-115
1761.	124a	Antigenic Index	120-135
1762.	124a	Antigenic Index	145-153
1763.	124a	Hydrophilicity	41-43

1764.	124a	Hydrophilicity	52-55
1765.	124a	Hydrophilicity	63-69
1766.	124a	Hydrophilicity	91-95
1767.	124a	Hydrophilicity	108-115
1768.	124a	Hydrophilicity	120-135
1769.	124a	Hydrophilicity	146-153

It will be understood that the invention is described above by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.

TABLE II

The present invention does not include within its scope proteins comprising any of the 45 protein sequences disclosed in Annex I. As stated above, if the length of any particular protein sequence disclosed in PCT/IB99/00103 is x amino acids, the antigenic fragment of the present invention has at most $x-1$ amino acids of that protein. For each of the 45 protein sequences given in Annex 1, the value of x is given in the following table:

SEQ ID NO:	x	SEQ ID NO:	x	SEQ ID NO:	x	SEQ ID NO:	x
2	245	26	571	50	185	74	150
4	591	28	710	52	166	76	255
6	592	30	710	54	326	78	255
8	164	32	62	56	356	80	172
10	321	34	86	58	284	82	242
12	321	36	92	60	1978	84	242
14	124	38	103	62	1532	86	183
16	124	40	85	64	593	88	155
18	173	42	78	66	129	90	153
20	640	44	78	68	319		
22	761	46	219	70	619		
24	111	48	212	72	595		

ANNEX I

COPY OF

INTERNATIONAL PATENT
APPLICATION

PCT/IB99/00103

MENINGOCOCCAL ANTIGENS

This invention relates to antigens from the bacterium *Neisseria meningitidis*.

BACKGROUND

Neisseria meningitidis is a non-motile, gram negative diplococcus human pathogen. It colonises the pharynx, causing meningitis and, occasionally, septicemia in the absence of meningitis. It is closely related to *N. gonorrhoeae*, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N. meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al.* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al.* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N. meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak

immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H. influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al.* (1996) *supra*; Costantino *et al.* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschroom (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal-transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of

further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the *N.meningitidis* amino acid sequences disclosed in the examples.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the *N.meningitidis* amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MSPRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the *N.meningitidis* amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure form (*ie.* substantially free from other *N.meningitidis* or host cell proteins)

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the *N.meningitidis* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the *N.meningitidis* nucleotide sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the *N.meningitidis* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1XSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the *N.meningitidis* sequences and, depending on the particular sequence, *n* is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*) but are preferably *N.meningitidis*, especially strain A, strain B or strain C.

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

Unlike the sequences disclosed in PCT/IB98/01665, the sequences disclosed in the present application are believed not to have any significant homologs in *N.gonorrhoeae*. Accordingly, the sequences of the present invention also find use in the preparation of reagents for distinguishing between *N.meningitidis* and *N.gonorrhoeae*.

25

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning: A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); *the Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice, Second Edition* (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9800760.2, 9819015.0 and 9822143.5 are incorporated herein.

Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature.

The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

10 An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

20 A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 30 5,753,233).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

15 Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

20 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host

range. Examples include the SV40 early gene enhancer. [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whiteley (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicon systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHERO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a

convenient restriction site for insertion of the heterologous gene or genes to be expressed, a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

- 5 After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit).
- 10 These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

- Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

- 20 Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

- 5 Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

- 10 Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedrin protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler). EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

- 15 DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Venheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Natl Acad. Sci. USA*, 82:8404; mouse IL-3, Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

- 25 A recombinant polypeptide or polypeptide may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus genome are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 µm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative

of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, eg. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as:

- US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry*, 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, J. *Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirtel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R. L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

- Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilimink and Dons, 1993, *Plant Mol. Biol. Repr.* 11(2):165-185.

- Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as T₁ sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance

toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

- The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

- A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.
- Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's spliceosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.* 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al.,

Nature, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browallia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop

simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (E. coli) [Raibaud et al. (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactase, lactase (*lac*) [Chang et al. (1977) *Nature* 198:1056], and maltose. Additional examples

include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g-lactamase (bla)* promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser), bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and TS [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *lac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Mann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Magai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *BioTechnology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masai *et al.* (1983), in: *Experimental Manipulation of Gene Expression*, Ghayeh *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline

phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

- 5 Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription.
- 10 Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

- 10 Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776; EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation.

Transformation procedures usually vary with the bacterial species to be transformed. See eg.

- 20 [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColEI-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS*

- Microbiol. Lett.* 44:173. *Lactobacillus*: [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]: [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III), Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, *Streptococcus*].

V. Yeast Expression

- Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

- Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

- In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GALL0*,

- OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 64 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henkoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*", in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Pulgaon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

- A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

- Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See eg. EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (eg. WO88/024066).

- Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in*

in vitro. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alpha factor. (eg. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Bolstein *et al.* (1979) *Gene* 8:17-24], pCUI1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci. USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and

usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the expression-construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

- Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *BioTechnology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

- Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See eg. [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; Candida]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenuhl]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *BioTechnology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach and Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody"

includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies. Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisseria proteins.

- Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by *in vitro* immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.
- Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the

immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAB. Further, one may combine various labels for desired effect. For example, MABs and avidin also require labels in the practice of this invention: thus, one might label a MAB with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAB labeled with HRP. Other permutations and possibilities will be readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or

combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to a pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals, in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hypodermis. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, poly(lactic acids), poly(glycolic acids), polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to:

(1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2%

Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphoryl lipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetyl-muramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (eg. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment

of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734).

5 Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Corneli (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses

eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698,

WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfield (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahrini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed

in Carter US Patent 4,797,368 and Muzycka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kolin WO94/288 57. Yet a further example of an AAV vector employable in this invention is SSV9A/FABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are disclosed in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/CP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Brakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-332), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositaries or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J. Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86; Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J. Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J. Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108; human immunodeficiency virus as described in EP-0386882 and in Buchsacher (1992) *J. Virol* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyuzlagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; ONyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curriel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials,

hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US 5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem* 262:4429-4432, insulin as described in Hueckel (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for

example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

- Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697, and WO91/14445; in EP-0524968, and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

- A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

- Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

- Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hypodermis. Dosage treatment may be a single dose schedule or a multiple dose schedule.

- Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate

precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

- In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

- One example are polypeptides which include, without limitation: asialoglycosaminoglycan (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of *Plasmodium falciparum* known as R11.

- B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

- Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

- The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-(2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectase (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dipoleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios.

Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Srittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B, and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (supra); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for

example, Atkinson (1986) *Ann Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

5 E. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polymethine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/EBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

20 The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene, Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when
25 combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods.

5 Antibodies to Neisserial proteins within biological samples, including for example, blood or serum samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

15 Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, etc.) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridization

20 "Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this binding include: the type and volume of solvent, reaction temperature, time of hybridization, agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [supra] Volume 2, chapter 9, pages 9.47 to 9.57.

25 "Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of
30

the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

- 5 Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10⁹ to 10⁴ g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10⁴ cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10⁴ cpm/µg, resulting in an exposure time of ~24 hours.

15 Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.41\%(G + C) - 0.6(\% \text{ formamide}) - 600/n - 1.5(\% \text{ mismatch}),$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- 25 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are

nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- 5 In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization 15 and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

- 20 The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

25 The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid

probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [e.g. see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [e.g. see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement)

to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-7 show biochemical data and sequence analysis pertaining to Examples 1, 2, 3, 7, 13, 16 and 19, respectively, with ORFs 40, 38, 44, 52, 114, 41 and 124. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (◊) shows preimmune data; a triangle (◻) shows GST control data; a circle (○) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophobicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes (Gao *et al.* (1988) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol suppl* 11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N. meningitidis*, along with their putative translation products. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein. It is believed at present that none of the DNA sequences described herein have significant homologs in *N. gonorrhoeae*.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N. meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- a corresponding gene and protein sequence identified in *N. meningitidis* (strain A)
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence homology between species and strains.

Proteins that are similar in sequence are generally similar in both structure and function, and the homology often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ-PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

Dots within nucleotide sequences (eg. position 288 in Example 12) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 589 in Example 12) represent ambiguities which arose during alignment of independent sequencing reactions (some

of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposito *et al.* [Critical evaluation of the hydrophobicity of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psorl.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (eg. fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N. meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C

- for 2 hours. Two phenol extractions (equilibrated to pH 8) and one CHCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

- Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

- The 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern), the 3' primers included a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *Bam*HI-*Xho*I or *Eco*RI-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

- 5'-end primer tail: CGCGGATCCCATATG (*Bam*HI-*Nde*I)
CGCGGATCCGCTAGC (*Bam*HI-*Nhe*I)
CCGGAATTCTAGCTAGC (*Eco*RI-*Nhe*I)
 3'-end primer tail: CCCGCTCGAG (*Xho*I)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridised to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

- $T_m = 4 (G+C) + 2 (A+T)$ (tail excluded)
 $T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N$ (whole primer)

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

- Table 1 shows the forward and reverse primers used for each amplification. Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH_4OH , and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

C) Amplification

- The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl_2), 2.5 units *Taq*I DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo *Taq* polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

- After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

- The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds 95°C	30 seconds 65-70°C	30-60 seconds 72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

D) Digestion of PCR fragments

10 The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-2)b+ and further expression of the protein as a C-terminus His-tag fusion
- *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

25 10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the

whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

5 The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).

F) Cloning

10 The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

15 In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

20 The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1µg) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1 µl of each construct was used to transform 30 µl of *E. coli* BL21 (pGEX vector), *E. coli* TOP 10 (pTRC vector) or *E. coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E. coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2 ml LB+Amp (100 µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20 ml of LB+Amp (100 µg/ml) in 100 ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1 mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2 mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1 ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20 ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600 ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₆₀₀ 0.8-1. Protein expression was induced with 0.2 mM IPTG followed by three hours incubation. The culture was centrifuged at 8000 rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5 ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40 W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150 µl Glutathione-Sepharose 4B resin (Pharmacia) (previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10 ml cold PBS for 10 minutes, resuspended in 1 ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2 ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700 µl cold Glutathione elution buffer

(10 mM reduced glutathione, 50 mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21 µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45.31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26 kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis

To analyse the solubility of the His-fusion expression products, pellets of 3 ml cultures were resuspended in buffer M1 [500 µl PBS pH 7.2], 25 µl lysozyme (10 mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40 W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8 M urea, 0.5 M NaCl, 20 mM imidazole and 0.1 M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6 M guanidinium-HCl, 0.5 M NaCl, 20 mM imidazole and 0.1 M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20 ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600 ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₆₀₀ 0.6-0.8. Protein expression was induced by addition of 1 mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000 rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5 ml of either (i) cold buffer A (300 mM NaCl, 50 mM phosphate buffer, 10 mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8 M, 10 mM Tris-HCl, 100 mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40 W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 3000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the OD₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

L) His-fusion large-scale purification

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded

onto a Ni-NTA superflow column (Qiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole.

After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORF 44, CD1 mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORF 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORF 38, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD_{600nm}. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum

5 diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₅₀ was followed. The ELISA was considered positive when OD₄₅₀ was 2.5 times the respective pre-immune sera.

O) FACSscan bacteria Binding Assay procedure.

- The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dragoon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₅₅₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₅₅₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerythrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACSscan tubes and read. The condition for FACSscan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by

centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₅₅₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was

washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₄₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

- 50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well.
- 5 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.
- 10

Table II gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 1>:

```

15 1 ..ACACTGTGCT TTGCAACGGT TCAGGCAAGT GCTACCAAT GAAGACAG
51 1 AAGAAGATT ATATTAGAC CCGGTACAC GCAGCTTGC GTGTGTGATA
101 GTCATCTGG ATAGARAGG CACGGAGAA AAGAANAAG TAGAAGAAA
151 TTGAGTATG GAGATATATT TCAACGAGAA AGAGTACTA ACACGACAA
201 AACTACACCT ACTGCTGAA AAGACCTCA AGAGTCTGA CCAAGTGTGG
251 RACTGAARA TATCGTTA GCCAAGCG CAAATAGTC ACATACAAA
301 GCGACACAA AGGCTTGAAT TTGCGAAG AAGGCTGG AAGCAAGGC
351 GACACACGG TTCACTTGA CCGTATTGG TCGAATTGA CCGATACGT
401 GCTGAATCC GGAGGACCA CAAAGTAA CACACACAC GTTACCGATG
451 ACATTAAG GGTGAACCC CGGTACAA AGTTCGATA CCTGGCTGG
501 CCGCGCATC TAGACACAG TCGAGTCTT GAGCGCAT ACAGAACAA
551 GCAGCTGTA TGTGAAGC AAGACACG GCAGAACAC CGAAGTTAA
601 ATCGGTGGA AGACTCTGT TATTAAGAA AAGAC...

```

30 This corresponds to the amino acid sequence <SEQ ID 2; ORF40>:

```

35 1 ..TLIFATVQAS ANQEEQEDL YLDPVQRTVA VLIWSDKEG TGEKEVEEN
51 1 SDVAVFNEK GYLTAAREITX KAGNLIKIKO NGNTYSYSLK KDLTSLVIG
101 TEKLSFANG NKNVITSDTK GLNFAKETAG NGDITVHLN GIGSTLIDL
151 LNTGATNTV NDNVTDDEKK RAASVDKVLN AGNIGKVPV GTTASNDVDF
201 VRTYDTVEFL SADTKITTVN VESKONGKKT EVKIGAKTSV IKENO...

```

Further work revealed the complete DNA sequence <SEQ ID 3>:

```

40 1 ATGACAAA TATACCGCAT CATTTGGAAT AGTGCCCTCA ATGCTGNGT
51 1 GCTGATATCC GAGTACACG GACACACAC CAACGCGCC TCCGCAACCG
101 TGAAGACCG CGTATGGG ACAGTGTGT TTGACCGGT TCAGGCAAGT
151 GCTACCATG AAGACACAA AGAAGTTA TATTAGACC CCGTACACG

```

```

201 CACTGTGCC GTGTGTAGT TCAATTCGA TAAAGAGCG ACGGAGAAA
251 AAGAAAATG AAGAANAAT TCAGTATGG CAGTATATT CAACGAGAA
301 GGAGTACTAA CAGCAGAGA ATCACCTCT ARAGCGGCG ACACCTGAA
351 ATCAACAAA CAGGACCAA ACTTACCTA CTCGCTGAA AAGACCTCA
401 CAGATCTAC CAGGTGTGA ATGAAAAT TATGTTTAG CGCAACGGC
451 AATAAGTCA ACATCAAGC CGACACAAA CGCTTGAAT TTGCAAGA
501 ACGGCTGGG ACAGCGGCG ACACCGGT TCATCTGAC GGTATGTGT
551 CGATTGNC CGATAGCTG CTGAATCCG GAGCACCAC AAGGTACCC
601 AAGCAACAG TTACGATGA CGAANAATA CTGGCGCAA CGGTAAAGA
651 GCTATTAAC GCTGGTGGG ACATTAAGG CGTTAAACG GGTACACAG
701 CTTCGATTA GTTGATTTT CTTGCGACT ACACACAGT CGAGTTCTG
751 AGCGGAGTA CGAACAAC GACTGTAT GTGGAAGA ACACACCGG
801 CAGAAACC GAATTAATA TCGTGGAA GACTTCTGT ATTAAGAAA
851 AAGAGGAG GTTGTGTTT GGTAAAGA AAGCGAGA TGGTCTTCT
901 ACAGACGAA CGAAGGCTT AGTACTGA AAGAAGTA TTGATCAGT
951 AARANGCT GTTGGGAA TGAACACAC ACCGCTAAT GTTCAACAG
1001 GTCAGCTGA CAGTTTGA ACCTGTAT CAGGCACAA TGTACCTTT
1051 GCTAGTGA AGGTACAC TCGGACTGA AGTAAGATG ATCAGGCAA
1101 CATCACTGT ATGTATGAT TAATGTGG CGATGCCCTA AAGTCAATC
1151 AGCTGAAA CAGCGTTGG AATTGGAT CCAAGCGGT TCGAGTTCT
1201 TCGGCAAG TCATCAGCG CANTTTTC CAGCANGG AAGATGGA
1251 TGAACCTTC ACATTAATG CCGCACAA CATCGAGAT ACCGCAACG
1301 GTAAATAT CGATCGCC ACTTCAGT CCGCAGTT TTCCAGCTT
1351 TCGTGGCG CGGGGCGA TGGCCACT TTGACGTGG ATGGGACCG
1401 ATTGANTGC GCGACAGA AGACACAA ACCGTCGG ATTACATG
1451 TCGCCCCGG GTTAAAGG GGGGATTA CAAAGTCGG ACATCTAAA
1501 GCGGTGGCG AATCTGAA CAACGCATC GAAATGGG ACGGCAACG
1551 GCGTGGCG ATCGCCAG CAGTGGAC CCGAGGCTG GTTCAGGCT
1601 ATTGCGCG CAGAGTAG ATGGCATG GCGGCGAC TTATCGGGC
1651 AAGCGGCT ACCTACCGT CTACTCAGT ATTCCGAG CGGAAATTG
1701 GATTACAA GCACGGCT CCGCAATC CGCGGCCAT TTCGTGCTT
1751 CCCATCTG CCGTATCAG TGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 4; ORF40-1>:

```

35 1 MKNYRIWV SALMANVVS ELIRNTRKA SATVKAVLA TLFATVQAS
51 1 ANNEQEDL YLDPVQRTVA VLIWSDKEG TGEKEVEEN SDVAVFNEK
101 GYLTAAREIT KAGNLIKIKO NGNTYSYSLK KDLTSLVIG TEKLSFANG
151 NKNVITSDTK GLNFAKETAG NGDITVHLN GIGSTLIDL LNTGATNTV
201 NDNVTDDEKK RAASVDKVLN AGNIGKVP GTTASNDVDF VRTYDTVEFL
251 SADTKITTVN VESKONGKKT EVKIGAKTSV IFAEGLKLT GKONGESS
301 TDEGLVTA KEVIDAVKA GPRKMTTAN GQTQADKE TVTSGTNTF
351 ASGKGTATV SKDDGNITV HYDWNVGDAL NYNQLONGM NLDSKAVAGS
401 SKGVISGNVS PSKRNDETV NINAGNIEI TRNKNIDIA TSMTPQESSV
451 SLGAGADAPT LSVGDALNV GSKDNKPV ITNAPGVRE GDTNVAOLK
501 GVAQNLNRI DNDVGNHAG IAQAIATAGL VQYILPKSH MAITGGTYRG
551 EAGYAIYSS ISDGNWIK GTASNSRGH FGASASVQW W*

```

Further work identified the corresponding gene in strain A of *N. meningitidis* <SEQ ID 5>:

```

50 1 ATGACAAA TATACCGCAT CATTTGGAAT AGTGCCCTCA ATGCTGNGT
51 1 CCGGTATCC GAGTACACG GACACACAC CAACGCGCC TCCGCAACCG
101 TGAAGACCG CGTATGGG ACAGTGTGT TTGCAACGGT TCAGGCAAGT
151 GACACCATG AAGATGAGA AGAAGATTA GAATCCGTAC AAGCTCTGT
201 CCGTAGGAC ATTCAGCA GATGAGAG GAGCGGCGA TTGGAAGCA
251 TATCATATC ATGACTTAC GACACANG AATTGTAGA CCCATACATA
301 GTATGTACC TCAACCGCG CGACACCTG AAAATCAAC AAGACACCA
351 TGAACACAC AATGCCAGTA GCTTACCTA CTCGCTGAA AAGACCTCA
401 CAGGCTGAT CAATGTGAN ACTGAAAT TATGTTTGG CGCAACCGG
451 AAGAAAGTA ACATTAAG GACACCAA GGTGTGAT TCAGCAAG
501 AAGCGTGG ACAGCGCG ACACCGGT TCATCTGAC GGTATCGTT
551 CGATTTCAC CGATACGCT GCGGTCTT CTGCTCTCA CGTATGCG
601 GGTAAACCAA GTACATTIA CACTCGTCA GCAAGTATA AGGATGTCTT

```


This encodes a protein having amino acid sequence <SEQ ID 6; ORF40a>;

The originally-identified partial strain B sequence (ORF40) shows 65.7% identity over a 254aa overlap with ORF40a:

STLDTLNTGATTNVTNDNVTDEKKRASVKDVLNAGNNIKGVKPGTTA--SDNVDFV
orff40.pep

The complete strain B sequence (ORF40-1) and ORF40a show 83.7% identity in 601 aa overlap

70	80	90	100	110
----	----	----	-----	-----

180	190	200	210	220	23
-----	-----	-----	-----	-----	----

0r140a KKGKNGSSTDEGEGLVTAKEVIDAVNKAGRHKTTANGQTGADKFETVSTGNTVT

orf40-1.pep SKGMDETVINAGNNEITRANGKIDATISMTPOFSSVSLGAGADAPTLISVGD-ALL

[illegible]

Score = 85 (37.6 bits), E=pect = 1.5e-116, Sum p(11) = 1.5e-116
Identities = 18/32 (56%), Positives = 20/32 (62%)

Query: 169 TNGDTTVMLAGTSLTDLGASASHVADN 200
T D +HLNGI SLTDL S + GN
Subject: 1469 TGDADNMLAGTSLTDLGASATNNGN 1500

Score = 92 (40.7 bits), E=pect = 1.5e-116, Sum p(11) = 1.5e-116
Identities = 16/19 (84%), Positives = 19/19 (100%)

RAAS+KDV LNAGWN++GVK
1509 RAASVKDVLNAGWNVRGVK 1527

Score = 90 (39.0 bits), Expect = 1.5e-116, Sum P(11) = 1.5e-116
Identities = 17/28 (60%), Positives = 20/28 (71%)

Query: 226 SIIGQSENVDVFRITDIVEFLSADTTT 253
S Q EN+DFV TYDTV+FS D TT

on homology with Hsf, it was predicted that this protein from *N. me*

es, could be useful antigens for vaccines or diagnostics.

20-1 (61kDa) was cloned in pET and pGex vectors and expressed in *E.*

The products of protein expression and purification were analyzed by SDS-PAGE. The results of affinity purification of the His-fusion protein, and Figs. 6 and 7 show the results of affinity purification of the His-fusion protein, and Figs. 8 and 9

of expression of the GST-fusion in *E.coli*. Purified His-fusion pro

and ELISA (positive result). These experiments confirm that ORF40-1 is

... and that it is a useful immunogen.

IE shows plots of hydrophilicity, antigenic index, and AMPHI region:

Following partial DNA sequence was identified in *N.meningitidis* <SEQ

1 ATGTTACGt TGACTGctt AGCCGTATGC ACCGCCCTCG CTTTGGGGCGC
51 GTGTTGGCCG CAAATTCCG TGTCTCCCCC TCAAGCCAAA C-ACACCCCG

101 TTTCGCGCG ACAACCGA GCGGGTCCG TTACGTCAG AACCGCGCG
151 GCGCAGTTT AAATACGCA AACCCGAA CGCATCGCG TTTACGATTT
201 GGGTATGTC GACACTTGA GCAACTGGG CGTGAAACC GGTTTGTCCG
251 TCGATATTA CCGCTTGG

301 CCTGCGGGCA CTTTGTTCGA GCCGGATTAC GAACGCTCA ACGCTTACAA
351 ACCGAGCTC ATCATCATCG GCAGCGCGC CgCCAGGCG TTTGACAAAT
401 TCAGCAAAAT CCCCCCGCG CAGCGCGCG CAGCGCGCG CAGCGCGCG

451. AAGGAAGTG CCAATGAGGC ATCGACGCTG GCGCAATCT TC..

This corresponds to the amino acid sequence <SEO ID 8: ORF38>:

100

1 MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR
51 GDVOIIPONE RIAYVDLGM DLTSKLGVT GLSDVKNRLP YLEEFYKTK
101 PAGTLEFPDY ETIMAYKPOL IIGSRBAKA FDLNLEIAPT IXHTADTANL
151 KESAKERIDA AATTA
151 KESAKERIDA AATTA

5 Further work revealed the complete nucleotide sequence <SEQ ID 9>:

1 ATGTACGTT TGACTGCTT AGCGTATGC ACCGGCTCG CTTCGGCCG
51 GTTTCGGCCG CAATATCCG ACTCTGCCCC ACAACGCAAA GAAACAGGGG
101 TTTTCGGCCG ACAACGCAAA GAGCGCTCG TTACGCTCAA AACCGGGG
151 GCGGATGCTC AATATCCGCA AAGCGGCGA CGATACGCG TTTCGATTT
201 GCGGATGCTC GACACTTGA GCAACGCGG CGTAAACG GGTTCGCTG
251 TCGATAAAA CCGCTGCGG TATTTAGAG AATATTTCA AACGCAAAA
301 CTTCGCGCA CTTCGCTGA GCGGATATC GAAACGCTCA ACCCTTAA
351 ACCGAGCTC ATCATCTCG GCGGCGCGC CGGCAAGGG TTGACAAAT
401 TGAAGCAAT CCGCGCGAC ATCGAATAA CCGCGATAC CGCAACCTC
451 AAGGAAAGT CCAAGAGCG CATGACGCG CTGGCGCAA TCTTCGCTA
501 ACAGCGGGA GCGGACGAG TGAAGCGGA AATCGCGCA TCTTCGCTA
551 CCGGAAACG TCGCGACAA GGTAAAGGA AAGTTTGT GATTTCGCT
601 AAGCGGCGA AGATGCTCG TTTCGGCGG TCTTCGCTG GATTTCGCT
651 GCTGACAAA GACATGCGG TTTCGGCTG TCTTCGCTG TGGGCGCTG
701 GAGGACGAG TCACTGCTT ACTTGAAT ACTTGAATG AATTAAGAG
751 GACTGCTGT TTGCTCTGA CCGAAGCGG GCGATCGCG AACAGCTTC
801 GCGGCGCAA GAGCTGCTG AATATTCCT GGTTCGCGA AACAGCTTC
851 GGAAGAGG ACAGCTGCT TACTCTGCT CTGAATCTA TTTCGCGAGC
901 GGTTCGCGG AAGCTGCT GATTCGAGC AACAGCTTC CCGAGCTTC
951 TACGCGCGA AATTA

This corresponds to the amino acid sequence <SEQ ID 10: ORF38-1>:

1 MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR
51 GDVOIIPONE RIAYVDLGM DLTSKLGVT GLSDVKNRLP YLEEFYKTK
101 PAGTLEFPDY ETIMAYKPOL IIGSRBAKA FDLNLEIAPT IXHTADTANL
151 KESAKERIDA LAOTFENAE ADKLKAEIDA SEAKTAQO GKGKLVLY
201 NGKRSAPFP SSSLGWLHK DGVPAVDEA IREGSHOPI SFYLYKENP
251 DMLFYDRAA AIGEGQAQK DVLDNPLVAE TTKMKGOV YLVEETLYAA
301 GGAOELMAS KOYADAPNA K*

Computer analysis of this amino acid sequence reveals a putative prokaryotic membrane lipoprotein lipid attachment site (underlined).

Further work identified the corresponding gene in strain A of *N. meningitidis* <SEQ ID 11>:

1 ATGTACGTT TGACTGCTT AGCGTATGC ACCGGCTCG CTTCGGCCG
51 GTTTCGGCCG CAATATCCG ACTCTGCCCC ACAACGCAAA GAAACAGGGG
101 TTTTCGGCCG ACAATCCGAA GCGCTGCG TTACGCTCAA AACCGGGG
151 GCGGATGCTC AATATCCGCA AAGCGGCGA CGTACGCG GTTTCGATTT
201 GCGGATGCTC GACACTTGA GCAACGCGG CGTAAACG GGTTCGCTG
251 TCGATAAAA CCGCTGCGG TATTTAGAG AATATTTCA AACGCAAAA
301 CCGTCCGCA CTTCGCTGA GCGGATATC GAAACGCTCA ACCCTTAA
351 ACCGAGCTC ATCATCTCG GCGGCGCGC CGGCAAGGG TTGACAAAT
401 TGAAGCAAT CCGCGCGAC ATCGAATAA CCGCGATAC CGCAACCTC
451 AAGGAAAGT CCAAGAGCG TATGACGCG CTGGCGCAA TCTTCGCTA
501 AAGGCGGAA GCGGACGAG TGAAGCGGA AATCGAGCG TCTTCGCTA
551 CCGGAAACG TCGCGACAA GGTAAAGGA AAGTTTGT GATTTCGCT
601 AAGCGGCGA AGATGCTCG TTTCGGCGG TCTTCGCTG GATTTCGCT
651 GCTGACAAA GACATGCGG TTTCGGCTG TGAAGAGC ATCAAGAG
701 GAGGACGAG TCACTGCTT ACTTGAAT ACTTGAATG AATTAAGAG
751 GACTGCTGT TTGCTCTGA CCGAAGCGG GCGATCGCG AACAGCTTC
801 GCGGCGCAA GAGCTGCTA ACAATCCCT GGTTCGCGA ACAACGCTT

851 GGAAGAGG ACAAGCTT TACTCTTC CTGAACCTA TTTCGAGC
901 GCGGCGCG AAGAGCTT GATTCGAGC AACAGCTTC CCGAGCTTC
951 TACGCGCGA AATTA

This encodes a protein having amino acid sequence <SEQ ID 12: ORF38a>:

1 MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR
51 GDVOIIPONE RIAYVDLGM DLTSKLGVT GLSDVKNRLP YLEEFYKTK
101 PAGTLEFPDY ETIMAYKPOL IIGSRBAKA FDLNLEIAPT IXHTADTANL
151 KESAKERIDA LAOTFENAE ADKLKAEIDA SEAKTAQO GKGKLVLY
201 NGKRSAPFP SSSLGWLHK DGVPAVDEA IREGSHOPI SFYLYKENP
251 DMLFYDRAA AIGEGQAQK DVLDNPLVAE TTKMKGOV YLVEETLYAA
301 GGAOELMAS KOYADAPNA K*

The originally-identified partial strain B sequence (ORF38) shows 95.2% identity over a 165aa overlap with ORF38a:

15	ORF38.pep	10	20	30	40	50	60
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						
	MLRLTALVNC TALALGACSP QNSDSAPQAK EQAVSAQTE GASVTVKTAR						

orf38-1 GGAQELLNASKQVADAFNAAK

Computer analysis of these sequences revealed the following:

Homology with a lipoprotein (lipo) of *C. jejuni* (accession number X82427)

ORF38 and lipo show 38% aa identity in 96 aa overlap:

5 orf38: 40 EGASVTVKARGDQVQIPONPERIAVYDLGHLDTLSLGLVKTGLS-VDRNRLPYLEEVKT 98
 EG S V K + G + P+NP ++ + DLG+LDT L + ++ V LP + FK
 Lipo: 51 EGDSELYKDSIGENKTPKPSKVYLDLADGILDTFALALNDKRVAGVPANLEPKYLOQFN 110

10 orf38: 99 TKPAGTLFEPDYETLNAYKPOLIIIGSRAAKAFDKL 134
 G + + D+E +NA KP LIII R + K +DKL
 Lipo: 111 KPSVGGVQVDFEATNALKPDIIIIIGROSKEFYDKL 146

Based on this analysis, it was predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

ORF38-1 (32kDa) was cloned in pET and pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure

2A shows the results of affinity purification of the His-fusion protein, and Figure 2B shows the results of expression of the GST-fusion in *E. coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 2C) and FACS analysis (Figure 2D). These experiments confirm that ORF38-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 2E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF38-1.

Example 3

The following *N. meningitidis* DNA sequence was identified <SEQ ID 13>:

25 1 ATGAACCTTC TGACCACCGC ATCTCTGTCT TCGGCATCG CGCTCAGCAG
 51 TATGGCTGCC GCGGTGGCA CGGACACCC CACTGTGCA AAAAACC
 101 TCAGCTAGCT CTGCGGACAA GGTAAAGAG TCAAGTAACT CTACGGCTTC
 151 AACAAACAGG GTCTGACAC ATACGCTTCC GCGGTATCA ACGGCAACG
 201 CGTGCAATG CCGTGTCAAT TGGCAAAAT CGCAATGTG GAACATCTT
 251 ACGGCAAGA AGCGGCTTAT GTTTGGGTA CCGGCTGAT GGATGGCAA
 301 TCTCACCACA AACAGCCCAT TATGATTACC GCACCTGACA ACCAATCGT
 351 CTTCAAGAC TGTTCCCCAC GTTAA

This corresponds to the amino acid sequence <SEQ ID 14; ORF44>:

35 1 MKLLTALLS SAIALSSMAA AAGTNNPTVA KKTVSVCQO GKVKVTYGF
 51 NKGLTTTAS AVINGKRVQV PYNLDKSDNV ETFGKEGGY VLGTGVHOGK
 101 SYRKQPIMIT APDQIVFKD CSPR*

Computer analysis of this amino acid sequence predicted the leader peptide shown underlined.

Further work identified the corresponding gene in strain A of *N. meningitidis* <SEQ ID 15>:

5 1 ATGAACCTTC TGACCACCGC ATCTCTGTCT TCGGCATCG CGCTCAGCAG
 51 TATGGCTGCT GCGGTGGCA CGGACACCC CACTGTGCC AAAAACC
 101 TCAGCTAGCT CTGCGGACAA GGTAAAGAG TCAAGTAACT CTACGGCTTT
 151 AACAAACAGG GTCTGACAC ATACGCTTCC GCGGTATCA ACGGCAACG
 201 TGTGCAATG CCGTGTCAAT TGGCAAAAT CGCAATGTG GAACATCTT
 251 ACGGCAAGA AGCGGCTTAT GTTTGGGTA CCGGCTGAT GGATGGCAA
 301 TCTCACCACA AACAGCCCAT TATGATTACC GCACCTGACA ACCAATCGT
 351 CTTCAAGAC TGTTCCCCAC GTTAA

10 This encodes a protein having amino acid sequence <SEQ ID 16; ORF44a>:

1 MKLLTALLS SAIALSSMAA AAGTNNPTVA KKTVSVCQO GKVKVTYGF
 51 NKGLTTTAS AVINGKRVQV PYNLDKSDNV ETFGKEGGY VLGTGVHOGK
 101 SYRKQPIMIT APDQIVFKD CSPR*

The strain B sequence (ORF44) shows 99.2% identity over a 124aa overlap with ORF44a:

15 orf44-pep MKLLTALLS SAIALSSMAA AAGTNNPTVA KKTVSVCQO GKVKVTYGF
 10 20 30 40 50 60
 orf44a MKLLTALLS SAIALSSMAA AAGTNNPTVA KKTVSVCQO GKVKVTYGF
 10 20 30 40 50 60
 orf44-pep AVINGKRVQV PYNLDKSDNV ETFGKEGGY VLGTGVHOGK
 70 80 90 100 110 120
 orf44a AVINGKRVQV PYNLDKSDNV ETFGKEGGY VLGTGVHOGK
 70 80 90 100 110 120
 orf44-pep CSFPR
 11111
 orf44a CSFPR

30 Computer analysis gave the following results:

Homology with the LecA adhesin of *Eikenella corrodens* (accession number D78153)

ORF44 and LecA protein show 45% aa identity in 91 aa overlap:

35 orf44 33 TVSVVCQO GKVKVTYGF NKGLTTTAS AVINGKRVQV PYNLDKSDNV ETFGKEGGYVL 92
 LecA 135 SVAVVCQO GRLN VNYRFSAGVPTSAELRVNRLPYNLSASDNVDVTF-SANGYRL 193
 orf44 93 GTGVHOGKSYRKQPIMITAPDQIVFKDCSP 123
 LecA 194 TTNANDSNYRSQDIIVSAPNGQLYKDCSP 224

40 Based on homology with the adhesin, it was predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

ORF44-1 (11.2kDa) was cloned in pET and pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the His-fusion protein, and Figure 3B shows the results of expression of the GST-fusion in *E. coli*. Purified His-fusion protein was used to

immunise mice, whose sera were used for ELISA, which gave positive results, and for a bactericidal assay (Figure 3C). These experiments confirm that ORF44-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 3D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF44-1.

Example 4

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 17>

```

1  ..GGACACCAAT TCMAACCAAC CTTTCCGGA GCGACATAC AGCAGAGGT
51  GGGTGAAMAA GCCGAGCCG ATCCGAATAT TATCTTAAA GGCATGGTA
101  ACCGATATCA AACGAGAGAA AGCTGTGAA CCACTCGAC CGTATGGCA
151  AAGCAGGCGG GAGCGGCGAG CAGGTTGAA AGCTGAGC TACGAGCTT
201  TGAGAGGCGG GCACTGCGTA AGCTAGCCG TCCGCGCGC TATATGCCG
251  ACATCCCGCA AGCAGACCTC AAACCGAAA TCMAACCAAC GGCATGGTA
301  CCGATATATG CCTATCTGAA ACAGCTTCA AGCTGAGC TACGAGCTT
351  GACCAAGTA CAGCTGCGT AGCACTATG GGAATATA CAGGAGGCC
401  TAACGCGAG CCGAGCCGCA ATTANCCAC TGGCGTTAC CGTGTCCAC
451  TCAGGCGGAG GACCGGAGC GGTATGGGA TTAAACGAG TGGCGCGCC
501  CGACACCAT GCACATTT...

```

This corresponds to the amino acid sequence <SEQ ID 18; ORF49>:

```

1  ..GTEFKTTLGG ADIAGVGEK ARADAKIILK GIVNIOTEE KLESSTWQ
51  KONGESTIVE TLKLPSTFEP ALPKLTPGG YIADIPKNI RTEIELKQ
101  PEYAIKQLQ TKDVMWNY QLAYDKWDYK QEGILTGAA IXLAVTVYT
151  SGAGTGAVLG LKRVAAATD AAF...

```

Further work revealed the complete nucleotide sequence <SEQ ID 19>:

```

25  1  ATGCACCTGC TGGCAGCCGA AGGCATTAC CACACCAAT TGAATGTCA
51  GAAAGTACC GATTTCATC GCATCAAAAT GGTGTAAGC AATTACACA
101  AAGAGAGCT GAACGAAGC AACTGCCG TACGGTTAT CGCCAAACA
151  GCCAAACCC GTTCGGCTG GGTATCCGA CTCGAAGCA CGAATTTAA
201  AAGCACCCTT TCCGAGCCG ACATACAGC AGGGTGGT GAAAGAGCC
251  GAGCGATGC GAAATTTATC CTAAAGGCA TCGTTAAGC CATCCAAAC
301  GAGGAAGC TGAATCCA CTCGACCTA TGGCAAGC AGCGGAGAG
351  CGGACGAGC GTTGAAGC TGAAGTACC GAGCTTTGA GGGCGGAGC
401  TGGCTAAGT GACCGTCC GCGGCTATA TCGCGAGC CCCCAGGAC
451  AAGCTCAAA CCGAATCGA AAGCTGGCC AAGCGCGC AATATGCTA
501  TCTGAACAG CTTGAGAGC TGAAGAGT GAACTGGAC CAATGACG
551  TGGCTTACA CAATGGAGC TATTAACAG AAGGCTTAC CGAGACGGA
601  GCGGCAATTA TCGCACTGC GTTACCTG TACCTTAC GCGAGAGAC
651  CATTGGCTC TTGGGCAAG CAGGCTTCG TATGTTAT CAACACAAA
701  GCGAATATG GTACACCTC GAAAGCTG GACGAAGCA GTTGGATGG
751  AATATGATG GTTGCCTGC CTACCGAG CAGAGCGAC AATATGCTG
801  CTTCGAGCT GAACATGTC AGGATAGC AATGATCA CAACCTGAC
851  GTCACTGCT GCAATGGGG CAGTGGCCA CTGATTTAA CCGCTTCAA
901  GCGGCGAGC CTGAAGACA ATCTGAAGC GAATATCTT GAGGCTTTG
951  TGAATACCT GATGAGAG GAGCAAGTA AATTAACAA GTTGGATGG
1001  CACTACATG CCTATAGAT TGGCCATCC ATAGCGGCT GTGGCGGAC
1051  GCGCGGAGT AAGGCAAGT GTCAAGATG TGGATGCT GCGGCGGCTG
1101  GTGAATCTT TGGCAAGC CTACTGAGC GAGAGAGC TGGCGGCTG
1151  AATGTGAAG ACAGGAGCA AATCATGCT AAGGAGAGC TGGAGAGC
1201  GCGGTTGCG GGGTTGAGT AGGGAGATG GAGTACGCG GGAATGCTG
1251

```

```

5  1301 CTGCTGAGC GGTAGCAAT AATCTTTAA ATGATGAGA GGAATCTTG
1351 TTGAGTGA AATATGCTT ATGATGAGA GGTAGGAGC GGAAGCTT
1401 TTGAGTGA AATATGCTT ATGATGAGA GGTAGGAGC GGAAGCTT
1451 GAGAAATCA ATTACCTAT AATTCGGA ACCTTTGA AGTGTTCG
1501 TTATATATA ACCTAGAA TGGCAATGA TATTTCTG TAGTAAAT
1551 ATGAGTACT GTAATTTCA CAATCAAA TATAGTGG GATCTGCG
1601 GTGGGTTT AATGTTTC CCTATGAT ATTAAAGC AGCATATG
1651 AATGATTTA GAATTTGTA TCMAATGAA GCTATGAG AATGATTT
1701 CCAAGCTTG GTAGTGA GGTGTGCTG TACTCTTG CTGACAGAG
1751 CTGCTTTTC GGTAGTGA ACATATCA AATCAATC TCTTTTAA
1801 GATTCAAAA TTATGGGA AATCGTTG GAAATGCTG TTGCTGAG
1851 AGTGAAGAA ACATATGA TAGTAAAT AATGATTT GATTAATTA
1901 TTAGTGAAC CATMAAAA TAG

```

This corresponds to the amino acid sequence <SEQ ID 20; ORF49-1>:

```

15  1  MQLAAGGIH QHOLWQST REIGIKVGS NYSKNELMET KLPVAVIOT
51  AKTRSGMDTV LEGTEKTL SGADIDGAG ERKADAKII LKGIYRIOT
101  EKLKSNSTV WQKAGSGST VETLKLPSE GGLPKLTPK GGIYDIPKG
151  NLTELEKLA KQPEYAVIQ LQYKOVWNN OVQLATKND YKQELTAG
201  AATLAAATV VTSGAGTAV LGLVGAAGA TGAATASLV QASVFINNK
251  GNITGTLKEI GRSSYKNI VAVATGAV KIGASALNV SDROMINLT
301  VNLAAGSAA LINTAVNGS LKDLKANIL AALVTATKE ASKRIQLO
351  HYLAHKIHA IAGCAAAAN KRCQOGAIG AVAGELIAGT LLDGIDPSL
401  NYKDRAKTA KAKLAAGA ALSKGVSFA ANAAVAVEN NSLMDIOLR
451  LSGNYALCHS AGNAGESICE YRPLGLPFT SVSGEHLRN KFGNHWK
501  LIINTRNQV YFVGWISF VSTKSNISV VSGWLVANS PHNYLKEHS
551  NDRNSNQK AVAEMISOTL VGESVGSIC LTRACSVSS TTSKSPFK
601  DSKITGELG GSGVAGVEK TITVINIKOI DRETSANIK

```

Computer analysis predicts a transmembrane domain and also indicates that ORF49 has no significant amino acid homology with known proteins. A corresponding ORF from *N. meningitidis* strain A was, however, identified:

ORF49 shows 86.1% identity over a 173aa overlap with an ORF (ORF49a) from strain A of *N. meningitidis*:

```

35  orf49.pep          10          20          30
orf49a  SKNELMETKLPVAVIOTWQKAGSGSTVETLKLPSEFSPALPKLTAAGYIADIPKNI
40          40          50          60          70          80          90
orf49.pep  GIVNIOTEEKLESSTWQKAGSGSTVETLKLPSEFSPALPKLTAAGYIADIPKNI
orf49a  GIVNIOTEEKLESSTWQKAGSGSTVETLKLPSEFSPALPKLTAAGYIADIPKNI
45          100         110         120         130         140         150
orf49.pep  KTELEKAKPEYAVIQLQYKOVWNNQVLAVDKWDYKQELTGAAIYALAVTVYT
orf49a  KTELEKAKPEYAVIQLQYKOVWNNQVLAVDKWDYKQELTGAAIYALAVTVYT
50          160         170         180         190         200         210
orf49.pep  SGAGTGAVLGLKRVAAATDAAF
orf49a  SGAGTGAVLGLKRVAAATDAAF

```


This corresponds to the amino acid sequence <SEQ ID 24: ORF50>:

```

1  ..RIVGRLISC AVTVESIT QGFVAFNSD KGVDAVGLA VLGTFVPHR
51  IGLRLIAS MLILFISNF STRLRASAK LSNALISGC SCLLFQSTIA
101  PTTAPLPFV A*

```

Computer analysis predicts two transmembrane domains and also indicates that ORF50 has no significant amino acid homology with known proteins.

Based on the presence of a putative transmembrane domain, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 6

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 25>

```

1  ..AGATTGACT TTAAGTGGT TATTCGGCG GATACAAAT ACCGCGGCTT
51  GTTTTGTGA GTATTGGTG TGTGGGGT GTTGACACTG TTTCGCTGA
101  TTAAGCTCT GTTTTTCGA TGTGGGGT ACNAGGTGCT GGTAATCGG
151  GGATCTCTA GTTTGATGT GGTGGGGT GCTTGTGTG TGTGTGCTT
201  GTTGAGATT GTTGGGGG GTTGGGGAC GATATGTTT GCACATACGA
251  CTTACAGTAT TGAATGGA TTGGGGGCG GTTGTTCGG GCATGCTGT
301  TCCCTGCTT TATCTATT CGACACAGA CAGATGGTG ATACGGTGG
351  TCGGGTGGG GAATGGAG ACATTCGCA TTCTTGACC GGGGACGCG
401  TACCTGGGT GTTGAATTT GCGTTTGTG TATCTTCTT GCGGTGATG
451  TGGATTACA GCTCAGCTT GACTGGGTG GTATTGGCTT CGTGT....
//
1451  .....
1501  .....
1551  .....
1601  CACCGGAGC GTCTGATTA TCGCCAGCG TCTGTGCTT GTTAAACGG
1651  CACCGCGAT CATTCGATG GATNAGGCA GGATTTGGA AGCGGGAAC
1701  CACGAGATG TCGTGGGCA CG..AACGA TATTACCGT ATCTGTATGA
TTACAGAC GGGTAG

```

This corresponds to the amino acid sequence <SEQ ID 26: ORF39>:

```

1  ..KDFWETPA VIKRRLFFE VLWVSVVLQ PALITPFFO VMDKVLVHR
51  GSTLDVSV ALVVSLEFI VLGGATLYL FHTTSRIDE LKARLPHLL
101  SLPLSTFHR RUGDTVARVA ELQDINFLT GQALTSVLQ ASSEFLPVAH
151  WYSSLTILW VLASL.....
//
501  .....
551  OQELNANKG YRVLVLDON G*

```

Further work revealed the complete nucleotide sequence <SEQ ID 27>:

```

1  ATGCTATAG TATCGGACC GCTCCCGCC CTTTCCGCC TCATATCTT
51  GCGCCATTC CACGCGATTG CCAGCAATC TCGCCATTA CAGCATGAT
101  TTTTACTTC CGCAGAGCG GATTMAATG AAGCGATTA CGGATTAGC
151  GCGAAATCT TGGATTTGA GCGCAAGGA GTCCGCGAC GATTTAAGC
201  TTGGCTAGT GCGACTTAC CCGCATGTG ATGGTGTAT GAGCGCAAC
251  ATTCAATTT GCGCAACGA GAGGTGAGG GTAGCATGC CCAATTTTG
301  ATACAGATT TGGTACGAA TAGCTGCG GTATGTCTT TTGCGCAAT
351  TTCTACAGA TATTCGGCA ACTGATGAT GTTGTCTTC GCGCTTGG
401  TATTGGGCG TTGGCAAGT TTGACTTAT CCGTGTAT TCGCGGGTA
451  ATCAATACG GCGGCTGTT TTTTACGTA TTGGTGGGT GGGTGGTGT
501  GCACTGTTT GCGCTGATTA CGCTCTGT TTTCACAGT GTATGACGA

```

This corresponds to the amino acid sequence <SEQ ID 28: ORF39-1>:

```

1  MEIVSAPDA LSAHLIANY HGIANPADI QHERTSASQ DUNGTOWLA
51  ANSLGUKRV VNOPIKRLM ATPLVWCO DGNHFLTK DEDEHAGTL
101  IODLVTKSA VLSFESNR YSGKLIVCA BASVLOSIAK PFTWRTFV
151  IYVRLEFEV LVSVVIGLE ALITPFFOV VMDKVLVHR GSTLDVSV
201  LLYVSLFEV LEGATTLFA HTTSRIDEL GARLPHLLS LPLSYFHR
251  VEDTVARVE LEQINFLG QALTSVLQ FEFILPVAH YSSLTILW
301  LASLPVAFW SAFISPLRT RLBDFRANA DMSFVESI TAVGTAKRA
351  VERPOTORD NOLAYVABG FRTYLVANG QCGVLOIKL VYVATLIGA
401  RLVESKLTV GOLATRMILS GOVAAPVIL AQUMDFQOV GISVATLGI
451  LMAFTENSS HLAIPDIRI IFEHVDFRV KADGRILLD INLIRAGEV
501  LGIVRSQSG KSTILKLVOR LVPEQGRVY VQDNDALAA PHLRNOGV
551  VLOENVILNR SIRINATLD TQPLENRIE AKLKAHEF IHELPQEGT
601  VVGEQAGLS GGORALIAI RALITNPRIL IFDELTALD YESRRAIQN
651  HOAICAMRV LIIRALSTV KTAHRIAND KQIVBAGTQ QELALPNQY
701  YRVLVDONG *

```

Computer analysis of this amino acid sequence gave the following results:

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF39 shows 100% identity over a 165aa overlap with an ORF (ORF39a) from strain A of *N. meningitidis*:

```

55      orf39 - pep      10      20      30
      KDFWETPAVIRKRLFFELVSVVLQ
      |||

```

	110	120	130	140	150	160
orf39a	AVLSFAESNRYSKGLILVAGSRVLSGLAKEDFTWIFPAVTKRRLFEVLVYSVLQGL					
orf39. pep	40	50	60	70	80	90
orf39a	FALITPFFQVMDKVLVHRCFSTLDVSVALLVWSLFEVLGGGLATYFAHTTSRDVE					
orf39. pep	170	180	190	200	210	220
orf39a	FALITPFFQVMDKVLVHRCFSTLDVSVALLVWSLFEVLGGGLATYFAHTTSRDVE					
orf39. pep	100	110	120	130	140	150
orf39a	LGARLFRHLISLPLSFEHRVGVGTVARVLELQIRNFQGLATSVLDLAFSIFLAVM					
orf39. pep	230	240	250	260	270	280
orf39a	LGARLFRHLISLPLSFEHRVGVGTVARVLELQIRNFQGLATSVLDLAFSIFLAVM					
orf39. pep	160	170	180	190	200	210
orf39a	WYYSSTLTWVVLASLXXXXXXXXXXXXXXXXXXXXXXXIXXXXXXXXXXXITIAHLSTV					
orf39. pep	290	300	310	320	330	340
orf39a	WYYSSTLTWVVLASLPAFYAFSFAFISPLIRLNMOKFARNQOSFELSVITAVGTVM					

ORF39-1 and ORF39a show 99.4% identity in 710 aa overlap:

orf39-1.pep	MSIVSAPILPALSLILAHYHGIANPADIQHEFCTSAQSDLNFTQWLAAKSLGKAKY MSIVSAPILPALSLILAHYHGIANPADIQHEFCTSAQSDLNFTQWLAAKSLGKAKY
orf39a	MSIVSAPILPALSLILAHYHGIANPADIQHEFCTSAQSDLNFTQWLAAKSLGKAKY MSIVSAPILPALSLILAHYHGIANPADIQHEFCTSAQSDLNFTQWLAAKSLGKAKY
orf39-1.pep	VQPIKRLRNATIPALVWCDDGNHFLIAKTDGGEHAQFLIQDLVTNKSAVLSFAEFSNR VQPIKRLRNATIPALVWCDDGNHFLIAKTDGGEHAQFLIQDLVTNKSAVLSFAEFSNR
orf39a	VQPIKRLRNATIPALVWCDDGNHFLIAKTDGGEHAQFLIQDLVTNKSAVLSFAEFSNR VQPIKRLRNATIPALVWCDDGNHFLIAKTDGGEHAQFLIQDLVTNKSAVLSFAEFSNR
orf39-1.pep	YSKGLLIVASRSVLSGLAKFDFTWFIPIAVIKXIRLFEEVLVSVYLQFALITPLFQV YSKGLLIVASRSVLSGLAKFDFTWFIPIAVIKXIRLFEEVLVSVYLQFALITPLFQV
orf39a	YSKGLLIVASRSVLSGLAKFDFTWFIPIAVIKXIRLFEEVLVSVYLQFALITPLFQV YSKGLLIVASRSVLSGLAKFDFTWFIPIAVIKXIRLFEEVLVSVYLQFALITPLFQV
orf39-1.pep	VHOKVLVHRGFSTLDVSVALLVWSLEIVLGGSLRTYLFANTSRIDVELGARLFRHLS VHOKVLVHRGFSTLDVSVALLVWSLEIVLGGSLRTYLFANTSRIDVELGARLFRHLS
orf39a	VHOKVLVHRGFSTLDVSVALLVWSLEIVLGGSLRTYLFANTSRIDVELGARLFRHLS VHOKVLVHRGFSTLDVSVALLVWSLEIVLGGSLRTYLFANTSRIDVELGARLFRHLS
orf39-1.pep	LPLSYFEHRRVGDTVARVREIQINFLTGQALTSVLDLAFSFFILAVHWYSSTLTWVY LPLSYFEHRRVGDTVARVREIQINFLTGQALTSVLDLAFSFFILAVHWYSSTLTWVY
orf39a	LPLSYFEHRRVGDTVARVREIQINFLTGQALTSVLDLAFSFFILAVHWYSSTLTWVY LPLSYFEHRRVGDTVARVREIQINFLTGQALTSVLDLAFSFFILAVHWYSSTLTWVY
orf39-1.pep	LASLPAYFMSAFISPIILATRLUNDKFEARNDAQSFVESITAVGTVKMAVEPQMTQRWD LASLPAYFMSAFISPIILATRLUNDKFEARNDAQSFVESITAVGTVKMAVEPQMTQRWD
orf39a	LASLPAYFMSAFISPIILATRLUNDKFEARNDAQSFVESITAVGTVKMAVEPQMTQRWD LASLPAYFMSAFISPIILATRLUNDKFEARNDAQSFVESITAVGTVKMAVEPQMTQRWD
orf39-1.pep	NQLAAVVASGFRVTKLAVVQGGVQLIQKLVTATLWIGARLVIESKLTVGQLIAFNHLS NQLAAVVASGFRVTKLAVVQGGVQLIQKLVTATLWIGARLVIESKLTVGQLIAFNHLS
orf39a	NQLAAVVASGFRVTKLAVVQGGVQLIQKLVTATLWIGARLVIESKLTVGQLIAFNHLS NQLAAVVASGFRVTKLAVVQGGVQLIQKLVTATLWIGARLVIESKLTVGQLIAFNHLS
orf39-1.pep	GOVAAPVIRLAQWQDFQVGSIVARIGDILNAPTENASSHLALPDIRGEITFEHVDFRY GOVAAPVIRLAQWQDFQVGSIVARIGDILNAPTENASSHLALPDIRGEITFEHVDFRY
orf39a	GOVAAPVIRLAQWQDFQVGSIVARIGDILNAPTENASSHLALPDIRGEITFEHVDFRY GOVAAPVIRLAQWQDFQVGSIVARIGDILNAPTENASSHLALPDIRGEITFEHVDFRY
orf39-1.pep	KADGRLLIQDLNLRIRAGEVLGIVGRSGSGKSTLTKLQRYLYPPEQGRVLVDGNLALAA KADGRLLIQDLNLRIRAGEVLGIVGRSGSGKSTLTKLQRYLYPPEQGRVLVDGNLALAA
orf39a	KADGRLLIQDLNLRIRAGEVLGIVGRSGSGKSTLTKLQRYLYPPEQGRVLVDGNLALAA KADGRLLIQDLNLRIRAGEVLGIVGRSGSGKSTLTKLQRYLYPPEQGRVLVDGNLALAA
orf39-1.pep	PAWLARQVGVVQLQENVLNFSIRDNIALTDTGHPLERIIEAAKLAGEHFHIELPEGYGT PAWLARQVGVVQLQENVLNFSIRDNIALTDTGHPLERIIEAAKLAGEHFHIELPEGYGT
orf39a	PAWLARQVGVVQLQENVLNFSIRDNIALTDTGHPLERIIEAAKLAGEHFHIELPEGYGT PAWLARQVGVVQLQENVLNFSIRDNIALTDTGHPLERIIEAAKLAGEHFHIELPEGYGT
orf39-1.pep	VVGQAGAGLGGGQRORIATARALINTPRILIFDTSALDYVESEIATIMONMAICANRTY VVGQAGAGLGGGQRORIATARALINTPRILIFDTSALDYVESEIATIMONMAICANRTY

This encodes a protein having amino acid sequence <SEO ID 30>:

1 MIVSRPUPA LSAIILIANY HGIANPADI QHEFCTSAQS DUNETQWLLA
51 ANSLGKAVY VPQPIKRAM ATIPALVWCZ DGNHILFAT VGSSEHAOYL
101 IDQUTWNSA VQSGEENFASR YGSKULIYAS RASVIGSLAK DFTFTWIPAV
151 IYKRLFEFV LVYVUQLQF ALITLPFOV QMOKIVVRG FSILDVUSYA
201 LWNLSFEIV LGLLPLUYA HTTSRIVDEL CARFELHLLS YSTLYFEHRR
251 VGDYAVARE LQIQUINFUC QALTSIVDEL FASFLIYVW YSTLTLYVW
301 LASIPAYW SAFIPISTI PLKDKFARNA ONQSELVESI TANGTVKAMA
351 VEPQUTQWD NQLAAYVASS FRYTKIYVQ QGVGQLIQKL VTATLVKICA

orf39a

VVGEAGLGGQRRIATRALITPRLIFDEATSALDYSERATMQNHQACANRTV

orf39-l pep

LITAHRLSTVTKTAHRIIANDKGRIVEAGTQQELLAKPNGYRYLYDLONGX

orf39a

LITAHRLSTVTKTAHRIIANDKGRIVEAGTQQELLAKPNGYRYLYDLONGX

The complete length ORF39a nucleotide sequence <SEQ ID 29> is:

[illegible]

401 BLVIESKLTV GQIAIENHLS GQVAAPVIRL AQLMDPQOV GISVARIGDI
431 LNAFTENASS HLALEPDIGE IFEVHDERF KADGELLIDQ LMLTIRAGEV
501 LGIVRSGSG KSTLTKLVOR LVYPAGOVVL VDGNDIADL PARYRAGOV
551 VLQENYLMR SIRONIATD TQPIERIE AAKKAGHER IMELPGIGT
601 VVEGOGAGIS GQONKJALA RALITNPRII IEDATISALD YESSRAIMON
651 HQALCNRTV LIIAHRLSTV KTAHRIIADH KGRIVEACTO OBLAARNGY
701 YRYLDQNG

ORF59a is homologous to a cytotoxin from *A. pleuropneumoniae*:

- 10 sp|E26760|RTIB_ACPRL RTX-I TOXIN DETERMINANT B (TOXIN RTX-I SECRETION ATP-BINDING PROTEIN) (APX-IB) (HLY-IB) (CYTOLYSIN IB) (CLY-IB)
>g1197137|p1c1|D43599 cytolysin IB - Actinobacillus pleuropneumoniae (serotype 9)
Score = 931 bits (2379), Expect = 0.0
Identities = 472/690 (68%), Positives = 540/690 (77%), Gaps = 3/690 (0%)
- 15 Query: 20 YGILANPADIQHEFCTASQSDMETQKXXXXXXXXXXXXVROPKIRLAHATLPALVWC 79
YH IA NP +++H+F + L+ T W V++ I RLA LPALVW
Sbjct: 20 YNIAVNPPELKHKEFDLEGK-IDLIAWLAIAKSLKAKQKALDIALATIALPALVWR 78
- 20 Query: 80 DQGNHFLAKTDGGGHAQYLLDITTKNSAVLSFAFSNRYSGKILVASRASVLSLA 139
+DG HFL K D E +YLI DL T+ +L AEF + Y GKILVASRAS++G LA
Sbjct: 79 EDGKHFLTKIDN--EAKRYLIDLETFNRLRLEQAEFSLQSKILVASRASVIGCLA 136
- 25 Query: 140 KEDFTWFIKAVIKYRAXXXXXXXXXXXXXXITPLFQVVMQKLVHREFGXXXXXXX 199
KEDFTWFIKAVIKYR+ ITPLEFQVVMQKLVHREF
Sbjct: 137 KEDFTWFIKAVIKYRKIFETLVSLFIQIFALITPLFQVVMQKLVHREFGSETLAVITV 196
- 30 Query: 200 XXXXXXXXFEIVAGLNTYLFPAHTSTRIDVELGARLFRHL+P-SYFEARRVGDVAVAR 259
FEIVL GLATY+FAH+TSRIDVELGARLFRHL+P-SYFEARRVGDVAVAR
Sbjct: 197 ALATVLFETVANGLTITIFRHSSTRIDVELGARLFRHLALPSTFEARRVGDVAVAR 256
- 35 Query: 260 ELEQIRNFTGQALTSVLDLAFSEFIFLAVMWYSSSTLTWVLSLPAVSAFISPLR 319
EL+QIRNFTGQALTSVLDL FSEIF AVHWYSS LT VLI SLR Y MS FISPILR
Sbjct: 257 ELQDIRNFTGQALTSVLDLAFSEFIFRVMWYSSFKTLVLVLSGLFTMGSSIFISPLR 316
- 40 Query: 320 TRLNDKFAHNDQSFVESITVGTVAHAAPVQMTORNDQALAAVVASGFVYTKLAV 379
RL++KFAF ADNOSFVES+TR+T+KA+AV PQMT ND QLA+YV++GFRVLT LA +
Sbjct: 317 RLDEKFAHNDQSFVESITVGTVAHAAPVQMTORNDQALAAVVASGFVYTKLAV 376
- 45 Query: 380 GQGGVQLIQKLVATLWIGARLVIESKLTVGQILAFNMLSGQVAAPVIRLAQLMDPQO 439
GQGGVQ IQR+V TLM+GA LVI L++GQILAFNMLSGOV APVIRLAQLMDPQO
Sbjct: 377 GQGGVQFIQKVVVITLWIGHVYISGDLISQILAFNMLSGQVAAPVIRLAQLMDPQO 436
- 50 Query: 440 VGISVARIGDILNAFTENASSHLAPDINGEITFEVHDERFKAQGLILDLMLIRAGE 499
VGISV RIGD+LA+PTE+ LALP+I+G+ITF ++ FRK D +IL D+NL I+ GE
Sbjct: 437 VGISVTRIGDVANSPTESYOGKALPEIKGIDITFRIRKRKPAADVILNDVLSIQOGE 496
- 55 Query: 500 VLGIIVRSGSGKSTLTKLVORLVYPAGOVVLVDGNDIADLAPARLVQVVLQENYLMR 559
V+GIVRSGSGKSTLTKL+OR Y+P G+YL+DG+DLAA P WLRQVGVVLQ+NYLMR
Sbjct: 497 VGIIVRSGSGKSTLTKLQIRTYIPENGQVLLDGHDLADLPWMLRQVGVVLQENYLMR 556
- 60 Query: 560 RSIRDNIALTDGM+P+I+I+AKKAGHEFI EL EGY T+VGEQAGISGGQONKJALA 619
RSIRDNIAL D GM+P+I+I+ AKKAGHEFI EL EGY T+VGEQAGISGGQONKJALA
Sbjct: 557 RSIRDNIALDPGHPREKIVHAAKAGHEFISLSENGINTVIGEGQAGISGGQONKJALA 616
- 65 Query: 620 ARAALITNPRIIIFDEATSAIDYESSRAIMONQALCNRTVLIIAHRLSTVKTARIRIIM 679
ARAL+ NP+I+IIFDEATSAIDYESS RAIMONQALCNRTVLIIAHRLSTVKTARIRIIM
Sbjct: 617 ARAALVNPRIIIFDEATSAIDYESSRAIMONQALCNRTVLIIAHRLSTVKTARIRIIM 676
- 70 Query: 680 DKGRIVEACTOELLAKRNGYRYLYDQNG 709
+KG+IVE G +ELIA PNG Y YL+ LQ+
Sbjct: 677 EKGQIVEGKHKELLADNGLYRYLDQNG 706

Homology with the HlyB leucotoxin secretion ATP-binding protein of *Haemophilus acetomycetemcomitans* (accession number X53955)

ORF59 and HlyB protein show 71% and 69% amino acid identity in 167 and 55 overlap at the N- and C-terminal regions, respectively:

- ORF59 1 KEDFTWFIKAVIKYRAXXXXXXXXXXXXXXITPLFQVVMQKLVHREFGXXXXXXX 60
KEDFTWFIKAVIKYR+ ITPLEFQVVMQKLVHREF
HlyB 137 KEDFTWFIKAVIKYRKIFETLVSLFIQIFALITPLFQVVMQKLVHREFGSETLAVITV 196
- 10 ORF59 61 XXXXXXXXFEIVAGLNTYLFPAHTSTRIDVELGARLFRHL+P-SYFEARRVGDVAVAR 120
FEIVLGLATY+FAH+TSRIDVELGARLFRHL+P-SYFEARRVGDVAVAR
HlyB 197 ALATVLFETVANGLTITIFRHSSTRIDVELGARLFRHLALPSTFEARRVGDVAVAR 256
- 15 ORF59 121 ELEQIRNFTGQALTSVLDLAFSEFIFLAVMWYSSSTLTWVLSL 167
EL+QIRNFTGQALTS+LDL FSEIF AVHWYSS LT VLI SL C
HlyB 257 ELQDIRNFTGQALTSVLDLAFSEFIFRVMWYSSFKTLVLVLSGLPC 303

//

- 20 ORF59 166 ICANRVTLLIAHRLSTVKTARIRIIMDKGRIVEACTOELLAKRNGYRYLYDQ 220
IC NRVTLLIAHRLSTV A RII MDK I+E G OELL + G Y YL+ LQ
HlyB 651 ICQNRVTLLIAHRLSTVKTARIRIIMDKGRIVEACTOELLAKRNGYRYLYDQ 705

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 7

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 31>

- 30 1 ATGAATACT TGATCCGAC CGCCTTACT GCAGTCGCAG CCGCCGCAT
51 CTACGCTGAC CAACGCCAAT CCGAAGCCGC AGTGAAGATC AAGGCTGAA
101 ACAGCCTGAC CGCTATGCGC TTACGCTGAC CCGAAGAA GCGAAGATT
151 GAGGCGTTGA AGCCGCAATK GAGCGCGAA ATCGA...

This corresponds to the amino acid sequence <SEQ ID 32; ORF52>:

- 1 MKVLIETALL AVAAGIYAC QPQSEAAVQV KAENSLTAMR LAVADKQAEI
51 DGLNAQIDAE INQRAEELK DYKIHQDAE VPELEK

Further work revealed the complete nucleotide sequence <SEQ ID 33>:

- 35 1 ATGAATACT TGATCCGAC CGCCTTACT GCAGTCGCAG CCGCCGCAT
51 CTACGCTGAC CAACGCCAAT CCGAAGCCGC AGTGAAGATC AAGGCTGAA
101 ACAGCCTGAC CGCTATGCGC TTACGCTGAC CCGAAGAA GCGAAGATT
151 GAGGCGTTGA AGCCGCAATK GAGCGCGAA ATCGAAGAA GCGAAGCCA
201 AGAATGAAA GACTACGAC GAAATACAGC GCAAGCGGAA GTGCCGAC
251 TCGAATAATG A

This corresponds to the amino acid sequence <SEQ ID 34; ORF52-1>:

- 1 MKVLIETALL AVAAGIYAC QPQSEAAVQV KAENSLTAMR LAVADKQAEI
51 DGLNAQIDAE INQRAEELK DYKIHQDAE VPELEK

Computer analysis of this amino acid sequence predicts a prokaryotic membrane lipoprotein lipid attachment site (underlined).

ORF52-1 (7kDa) was cloned in the pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion. Figure 4B shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF52-1.

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 8

10 The following DNA sequence was identified in *N. meningitidis* <SEQ ID 35>

1 ATGCTTATCG GAATATTAAT CGCATCAAGC AAGCATGCTC TTGCTATTAC
51 TCTATGTTGA ATCCCTGCTC TCCATGCATC CAGTTCGGTA TCGCGTTAGG
101 CAATACGGAA TAARATCTGC TGTTCTGCTT TGCTATATTT TGCCAAATG
151 TTATGTTT CTTAGAGGAC AGCTTGCTTA CCGGCTTCGC CTTTGCACAA
201 GCGCCCCACA GCGCTTCCC AAGCTTGCC TACCGTTACC GCACCGGTG
251 CGATTCGCGC GCGCGCTTCG GCAGCTGA

This corresponds to the amino acid sequence <SEQ ID 36; ORF56>:

1 MVIGILLASS KVALVITLL NPVFHASSCV SRKAIRNKIC CSALAKENKI
51 FVLSGNACL AFAFDNAPT GASQALPTV APVAIPAPAS AA*

20 Further work revealed the complete nucleotide sequence <SEQ ID 37>:

1 ATGCTTATCG CAGGTTTGAT GGTITTTCCG TTAATGGTTA TCGGAATATT
51 ACTTGCATCA AGCAGCGCTG CTCCTTTCTC TACTTATTG TTAATCCCG
101 TCTTCCATGC ATCCAGTTGC GTATCGGCTT GGCAATATAC GAATAAAATC
151 TGCTGTTCTG CTTTGGCTAA ATTGSCAAA TTGTTATTG TTCTTTAGG
201 AGCAGTTGCG TTAGCGGCTT TCGCTTCCA CACAGCCCCC ACAGGCGCTT
251 CCCAAGCGTT GCCTACCGTT ACCGACCCCG TGCGGATTCC GCGCGCCGCT
301 TCGCAGCGCT GA

This corresponds to the amino acid sequence <SEQ ID 38; ORF56-1>:

1 MACTGLMVEP LMVIGILLAS SKFAPELTLL LNPVFHASSC VSRWAIRNKI
51 CCSALAKENK LTVLSGNAAC LAFAFDNAP TGASQALPTV APVAIPAPA
101 SAA*

Computer analysis of this amino acid sequence predicts a leader peptide (underlined) and suggests that ORF56 might be a membrane or periplasmic protein.

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 9

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 39>

1 ATGTCAGTA TTTAAATGT GTTCTTCAT TGTATCTGG CTTGTGTAGT
51 CTTGTGTAG AGCCTACTA TATTGTGAT CTTGTCTCT TTTACTTAT
101 TGTATCTTC TTATCTGCT GTTTTAAAGA TTTCTTTTC TTTTCTTAA
151 CACAGATTT CACTCGGCTG TCCAGGCTG GAGTCAAAAT GGCATGACCC
201 TTGCGTCAC TGGCTACGG CCACTCTGCG TATTCGCGG CCTCAGCCTC
251 CAGGG...

This corresponds to the amino acid sequence <SEQ ID 40; ORF63>:

1 MESTINVELH CILACVYSGE TPTIEGLAL FYLLYLSYLA VFRIFFSFFL
51 DRVLSRSPRL ECKWHOPLAH WUTATSAILP POPPG...

Computer analysis of this amino acid sequence predicts a transmembrane region.

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 10

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 41>

1 ..GTGCGGCGT GGTGTGTTT TTGTTGCGAC GGTITGAAT ACCCGTTGTT
51 GCTTGTGAT GCGGATATGT TGCTGTACCG GTTGTGGGC GCGCGGAAA
101 TCGAATGCGG CCGTTCCTT GTGCGCGGA TGACGATTT GCAGCATTT
151 TTGCGGCGGA TGGGACCGT GTGCGCTGG GTGCGGCTGA TTGGGCTATA
201 CCGTATGAT GAATGATGAAA AAGACGGAAG ATATTGA

This corresponds to the amino acid sequence <SEQ ID 42; ORF69>:

1 ..VRTWLVFWLQ RLKYPILLMI ADMLYRLLG GAIEGCRCP VPPTMDWQHE
51 LPAMGTVSAW VAVIWAYLMI ESEKNRY*

Computer analysis of this amino acid sequence predicts a transmembrane region.

A corresponding ORF from strain A of *N. meningitidis* was also identified:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF69 shows 96.2% identity over a 78aa overlap with an ORF (ORF69a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf69 .pep	VRTWLVFWLQRLKYPILLMIADMLYRLLGGAIEGCRCPVPPTMDWQHFLPAMGTVSAW					
orf69a	VRTWLVFWLQRLKYPILLMIADMLYRLLGGAIEGCRCPVPPTMDWQHFLPAMGTVSAW					
	10	20	30	40	50	60
orf69 .pep	VAVIWAYLMI	ESEKNRYK				
	70	79				

VAVIWAYLHIESEKNGRXX
70

The ORF69a nucleotide sequence <SEQ ID 43> is

10 This encodes a protein having amino acid sequence <SEQ ID 44>:

10 This encodes a protein having amino acid sequence <SEQ ID 44>

1 VRTWLFWLQ RLRPLLCI ADMLYRLLG GAEIEGRCR VPPTMDQHF
51 LPTMGTVAAW VAVIMAYLMI ESEKNGRY*

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 11

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 45>

1
5
10
15
20
25
30
35
40
45
50
55
60
65

ATTGTTCAAA ATTGATTGT GAGGCGATTC GTGGATGAGG TCTCTCCCGT
GCTCGGCTCC ATTACGCTCT CGCAGGTGATC CCGGCGGTAT ACGGAGGATG
ACTGGGAGAA CAACTACTGC GAACTAATAG CGAGGAGCTT ACAGACACCCC
CTGGCCCAATA TCGATTGTGT CGGACAACTAT GTCTATACC CTCTGACATT
GATTCTGAG CCGTCTCTGT TGGGCTGGAG GGTGCGGAT CTTCATCATTT
CGGGCAACT CGGCACCCG CCGCTTGCTT GGCCTTGTGCT TGCCGCTATTC
GGCCCGCTGT CCAATCTAGC GATGGCTGT* GTTGGGAGG TGTTTATGAG
GCTACTGCGG TATTCGCGG GGGCGATCA GATCGCGTGG GCTCAATATG
CAAACTAGG TATTCGATC AATCGATTC TGTTGCGCT CAATCATATC
CCCATCTTCG CTTGGAGACG CGCATTTTC ATCTACACTT TCGTCTGGAG
GAAATATCTG CAAAGCTTCC GCAAAATAG ACTTATAGG AGCTAGATAT
TCTCATGCT GATCTGACC GGGCTTTGG GTTCGTTTAT WGCACCATTT
GTGGGATTC GTGATTCCT TTGTGCAAT GTACGCTTGA CTGGCTTTCA
GAGCGCATTA

This corresponds to the amino acid sequence <SEQ ID 46; ORF77>:

1 MEQNEFLAGE LAALVPLPS ITSXVAGY TARKWQNTA EYQGLTIAN
51 LPHLDVGTI IYPLLTMTF PFLGWARPI PDSNRNRP RLAMCVAAS
35 101 GPNSNARV LAGVULVTR YGAGVQRP ADMANGILI NALPALNIII
151 PLEPMGAGF IOTIFSAKYS QAFRIEYEG TWILLDDMTI XVLGFTAPI
201 XAKKDCXAD VALIGFOTA*

Further work revealed the complete nucleotide sequence <SEQ ID 47>

40
1 ATGTTTCAC ATTATGATTT GGGGGTGTT TCGATTGAGG TCTGTCGGCT
51 GCTGCTCTC ATTACCGTAC GGGAGGTGAG CGGCGGGTAT ACGAGGACCG
101 ACTGGGGAGCA CAACTATGCC GAACTATGCG GCGAGTGATC ACTGAGACCCC
151 CTGGCCCAAC TCGATTTGGT GCGGCAATAC ATCTGACGCC TGATTTATCAT
201 GATGTTCCAG CCGTTCCTGT TCGGGTGGGG GCGTCGATAT CCGATCATTT
251 CGGCGCACTT CCGCAACCGG CGCTTGTCT GGGCATTTGCT TGGCCGCTCT
301 GCTGACCTCTT CCAATCTGCG GATGGGTGTT TGTGGGGAGG TGGTTTATGG
351 GCTACTACCG TATGTGGGGG GGGCGATACA GATGCCGTTG GCTAAATATG
401 CAAATACGCG TATTCGATTC ATGTCGATTC TGTTCGGGCT CAAATCATGC

451 CCCATCTG CTTGGACGG CGGATTTC ATGCACACT TCCGTGCGC
501 GAAATATCG CAAGCTTTC GCAAATGCA ACCTTAAGG ACGTGAATT
551 TCCACTGCT GATGCTGAC GGGGTTTGG GTGGTTTAT TGCACGAATT
601 GTGGGCTGG TATTGCGT TTGCGAGT TTGCTCTGA

5 This corresponds to the amino acid sequence <SEQ ID 48; ORF77-1>

1 MEONFODJGE LLAATPEYDUS ITVEVARGO TARVIGONTA EOYGLITLANE
51 LEHDIJLHET IYPTLITMET PELGGAARDI PIDSNNRNP BLAKCTAASIS
101 GEDSNAAY IAGVYUWLP YGGAQYOML AOMANVCLIL NALLTAMUILL
151 PLEHMOGGE ITTELAKTYS QAFKRIEYQ TWILLILHET GULGATFAP
201 VALUAFYOM FY*

Computer analysis of this amino acid sequence reveals a putative leader sequence and several transmembrane domains.

A corresponding ORF from strain A of *N. meningitidis* was also identified.

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF 11 shows 96.5% identity over a 173aa overlap with an ORF (ORF77a) from strain A of *N. meningitidis*:

10 20 30 40 50 60
 MFQNFDSGFVLLAVLPVLPSTSVARGTARKWNGDNTAOGRLTJLNPIDLVGTT
 or f77, pep
 or f7a
 RGTARYWGDNTPAOGRLTNPIDLVGTT
 20

25
or17a
70 80 90 100 110 120
..TYP.LT.LM.T.E.F.L.E.G.A.R.P.I.P.D.S.N.F.R.N.R.I.A.R.C.V.A.S.P.S.N.I.A.N.A.V.L.G.V.L.Y.T.F.T.
I.P.L.T.M.T.E.F.L.E.G.A.R.P.I.P.D.S.N.F.R.N.R.I.A.R.C.V.A.S.P.S.N.I.A.N.A.V.L.G.V.L.Y.T.F.T.
40 50 60 70
or17a
..TYP.LT.LM.T.E.F.L.E.G.A.R.P.I.P.D.S.N.F.R.N.R.I.A.R.C.V.A.S.P.S.N.I.A.N.A.V.L.G.V.L.Y.T.F.T.

	140	150	160	170	180
30	YGGAGVOMHLOMAN	GLILINLALP	LEPILEPMGGG	ITDFP	SAKSOARKEIP
or177_psp					
or177s	YGGAGVOMHLOMAN	GLILINLALP	LEPILEPMGGG	ITDFP	SAKSOARKEIP
100					
110	YGGAGVOMHLOMAN	GLILINLALP	LEPILEPMGGG	ITDFP	SAKSOARKEIP
120					
130	YGGAGVOMHLOMAN	GLILINLALP	LEPILEPMGGG	ITDFP	SAKSOARKEIP
140					
150	YGGAGVOMHLOMAN	GLILINLALP	LEPILEPMGGG	ITDFP	SAKSOARKEIP
160					

orf77a	orf77b	orf77c	orf77d	orf77e	orf77f	orf77g	orf77h	orf77i	orf77j	orf77k	orf77l	orf77m	orf77n	orf77o	orf77p	orf77q	orf77r	orf77s	orf77t	orf77u	orf77v	orf77w	orf77x	orf77y	orf77z																																																											
160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660	670	680	690	700	710	720	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870	880	890	900	910	920	930	940	950	960	970	980	990	1000

40 ORF77-1 and ORF77a show 96.8% identity in 185 aa overlap

	10	20	30	40	50	60
orf77-1.psp	MKQNFQGVFLLAFLPVL	LSLITVEVARGGTRRG	SGMGTAAOYGLTLNPL	PIDLVGT		
orf77a						
			RGTRRWGMDTAQIGRLTLNPL	PIDLVGT		
		10	20	30		
orf77-1.psp	70	80	90	100	110	120
	IYPLILMTPELFGARPI	PDISRNFRBLAECV	ASQSPSLNMAALV	GGVLYLTP		

or f77a 1 VPLLTLHTPELEGWAPIDISRNFRPLAWCAVAGSGLSNLAHVLWGVLVLP 40 50 60 70 80 90 180

5 or f77-1-pep 1 YVGAYQHPLAQHANYGILINAILFALNIPIDWDGDFIDITFSLSKYSAFRKIETPYG 130 140 150 160 170 180

or f77a 1 YVGAYQHPLAQHANYGILINAILFALNIPIDWDGDFIDITFSLSKYSAFRKIETPYG 100 110 120 130 140 150 160 170 180

10 or f77-1-pep 1 TWIILLMLTGVLFARFARFVRLVIRFQVFX 190 200 210

or f77a 1 TWIILLMLTGVLFARFARFVRLVIRFQVFX 160 170 180

15

A partial ORF77a nucleotide sequence <SEQ ID 49> was identified:

1 ..CGCGGCTATA CAGGCGGCTA CTGGGGTGAC AACACTGCGG AACAACTACGG 51

51 CAGGCTGACA CTGACACCCG TGCCCCATAT CATTGCTG GGCACATCA 101

101 TCGTACCGCT GCTTACTTTC ATGTTTACCG CTTCTCTGTT CCGCTGGCGG 151

151 GCTGCGATTG CTATGATTC GCGCACTTC CCGAACCCG CCGTTCGCTG 201

201 GCGTTCGCTT GCGGCTCGG CCGGCTGTC GATCTGCGG ATGCTGTTTC 251

251 TGTGGGGCTT GGTTTTGTG CTGACTCCGT ATCTCGGTGG GGGTATACG 301

301 ATGCGCTTGG CACAAATGCG AACACTACNN ATTCATGACA ATGCGATTC 351

351 TGGGCGCTC AACATATCC CCATCTGCG TTGGGACCG GGCATTTC 401

401 TCGACACCTT CCGTGGCNC AANTATCGC AAGCTTCCG CAATAATCGA 451

451 CTTATGGA GCGGATAT CCNGCTGCTT ATCTGACCG GGGTTTGGG 501

501 TGGTWTAT GCACCATTTG TCGACGTGTT GATGCGTTT GTGCAGATG 551

551 TCGTCTGA

This encodes a protein having amino acid sequence <SEQ ID 50>:

30 1 ..RGVARYGND NTAQVGRIT LAPLEHDLV GTLIPLTL HTPELFQWA 51

51 RPIPDARNF RNPRLMRCV AASGLSNLA MAYLGVVLA LTPYVGAYQ 101

101 MPLAQHANYX ILINAILXAL NIPILPMDG GFIDITFSLA KXSAFRKIE 151

151 PYGTWIXLL MLTGVLAGXI APIVOLVIAF VQVFX

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 12

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 51>

1 ATGACCTGA TTACAGTTA CATCATCGT CAATGCGG TTATGCGGT 51

51 TTACGCTC CTGCTCTCC TCGTTTGTG CAGCTTTT GAAATCTGT 101

101 ACGAACCGG CAACCTCGG AAGCGAGT AGGCATATG GGAATGCTG 151

151 GGCTACCG CCTCBAAT GCCGCCCG GCTACGAG TGATCCCT 201

201 GCGGCTCT ATGCGGAC TGTCTCTCT CAGCAGCT CCGCGCGCA 251

251 GCGAAGTAC GGTCACTAA GCGAGCGCA TGACACCA AAGCTGCTG 301

301 TTGATCTGT CCGAGTTCG TTTATTTT GCTATGCCA CCGTGGCT 351

351 CCGGCAATG GTGCGGCA CACTGAGCA AAGCCGCA ACATCAAG 401

401 CCGCGGCT ATCGACCA ATCAGACCG GCAATGCT CTTTGGCT 451

451 AAGAAAAA ACAGCTGAT CAATGTGCG GAAATGCT CCGACAT..

This corresponds to the amino acid sequence <SEQ ID 52> ORF112>:

1 MNLISRIIR QHVMVAVL LAFALYSEF EILYETGNLG KGSYGIWHL

51 GYTALKWPAR AYELIPLAVL IGLVLSQL AAGSELTIVK ASGMTKKLL 101

101 LILSQGFIF AIATVALGEW VAPTLSQAE NIKAAINGK ISTGNTGLML 151

151 KENSVINVR EHLDPH...

Further work revealed further partial nucleotide sequence <SEQ ID 53>:

5 1 ATGACCTGA TTACAGTTA CATCATCGT CAATGCGG TTATGCGGT 51

51 TTACGCTC CTGCTCTCC TCGTTTGTG CAGCTTTT GAAATCTGT 101

101 ACGAACCGG CAACCTCGG AAGCGAGT AGGCATATG GGAATGCTG 151

151 GGCTACCG CCTCBAAT GCCGCCCG GCTACGAG TGATCCCT 201

201 GCGGCTCT ATGCGGAC TGTCTCTCT CAGCAGCT GCGCGCGCA 251

251 GCGAAGTAC GGTCACTAA GCGAGCGCA TGACACCA AAGCTGCTG 301

301 TTGATCTGT CCGAGTTCG TTTATTTT GCTATGCCA CCGTGGCT 351

351 CCGGCAATG GTGCGGCA CACTGAGCA AAGCCGCA ACATCAAG 401

401 CCGCGGCT ATCGACCA ATCAGACCG GCAATGCT CTTTGGCT 451

451 AAGAAAAA ACAGCTGAT CAATGTGCG GAAATGTC CCGACATAC 501

501 GCTTTGGCG ATCAAAATTT GCGCGGCA CATAAAGG GAAATGCGAG 551

551 AGCGATGA AGCGATTC CCGTTTGA ACAGCAGG CAGTTGGCAG 601

601 TGTAAACA TCGCGCGG CAGCTTGGC GAAGACAA TCGAGTCTC 651

651 TATTGGCT GAAGAAACT GCGGATTC CGTCAACG ACCTGATG 701

701 ACATTTGCT CTTCAACCC GACCAATTC CGTGGCGA ACTGACCC 751

751 TACATCGCG ACCTCAGAA CACACCAA ACACCCGGA TCTAGCCAT 801

801 CCGATGCT GCGAATGG TTACCCCG CCGAGCTG GTGATGGC 851

851 TCGTGGCT TCGCTTTAC CCGCAACCA CCGCGCAG CAATATGGC 901

901 TTAAGACT TCGCGGCT CTTGTCGGA TTGCTGTC ACCTTGGCG 951

951 ACGGCTCTT GGGTTTACA GCAACTCG...

25 This corresponds to the amino acid sequence <SEQ ID 54> ORF112-1>:

1 MNLISRIIR QHVMVAVL LAFALYSEF EILYETGNLG KGSYGIWHL 51

51 GYTALKWPAR AYELIPLAVL IGLVLSQL AAGSELTIVK ASGMTKKLL 101

101 LILSQGFIF AIATVALGEW VAPTLSQAE NIKAAINGK ISTGNTGLML 151

151 KENSVINVR EHLDPHLLG IKTWANDKN ELAEAEVDS AVLSDGSQ 201

201 LKNIERTSLG EDKVEVSTAA ENNPISYVR NIMDLVLRP DQMSYGLIT 251

251 YRHJONNSQ NTRIVAIWV RKLVPAAW VMLVAFAFT POTTGHNCG 301

301 LLEGGICGX LLEHGLGELF GFTSQL...

Computer analysis of this amino acid sequence predicts two transmembrane domains.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.meningitidis*:

	10	20	30	40	50	60
or f112-pep	MNLISRIIRQHVAVVALLAFALYSEF	EILYETGNLGKSGYGIWHLG	YTALKWPAR	AYELIPLAVL	IGLVLSQL	AAGSELTIVK
or f112a						
	10	20	30	40	50	60
or f112-pep	AYELIPLAVL	IGLVLSQL	AAGSELTIVK	ASGMTKKLL	LILSQGFIF	AIATVALGEW
or f112a						
	70	80	90	100	110	120
or f112-pep	AYELIPLAVL	IGLVLSQL	AAGSELTIVK	ASGMTKKLL	LILSQGFIF	AIATVALGEW
or f112a						
	130	140	150	160	170	180
or f112-pep	AYELIPLAVL	IGLVLSQL	AAGSELTIVK	ASGMTKKLL	LILSQGFIF	AIATVALGEW
or f112a						

orf112a-pep VAPLTSQKAEINIAAINGKISTGNTGLMEKNVINYREMLPDR
 orf112a VAPLTSQKAEINIAAINGKISTGNTGLMEKNVINYREMLPDRILTLGKIMANDRN
 130 140 150 160 170 180
 orf112a ELAEVADSAVNSDGSQMLKNIRSTLGEDKVEYSIAEXWPIYSKRNIMLDVLYRP
 190 200 210 220 230 240

A partial ORF112a nucleotide sequence <SEQ ID 55> was identified:

1 ATGACCTGA TTACACCTTA CATCATCCGT CAATGGGCG TTATGGCGGT
 51 TTACGGCTC CTTCGCTTC TCCTTTGTA CAGCTTTT GAAATCCTT
 101 AGAAGACGG CAACCTGGC AAGGCAATT AGGCATATG GGAATGNTG
 151 GGNATACCG CCTCTAAAT GNCGCCCGC GCTCAGAC TTATGCCCT
 201 GCGCTCCTT ATGGCGGAC TGGTCTCT CAGCGACAC TGGCGGGGA
 251 GCGACTGAT CGTATCAAA GCGACGGCA TGAACCCA AAGCTGCTG
 301 TTGATCTGT CCGATTCGG TTATATTTT GCTATTCGA CCGTGGCT
 351 CCGGATATG GTTGGCCCA CACTAGCCA AAGAGCCGA AACATCAAG
 401 CCGGGCCAT CAGCGCAA ATCAGTACG GCATACCGC CTTTGGCTG
 451 AAGAAAAA ACAGCATTA CATGTGCGC GAAATTTGC CCGACCATC
 501 CTGTGGGCG ATTAAATCT GGGCCCGCA GATTAAC GAACTGGAC
 551 AGCAGTGA ACCGATTC GCGGTTTGA ACAGGACG CAGTGGCAG
 601 TTGAAMCA TCGCGCGAG CAGCTTGGC GAGACAAAG TGAAGTCTC
 651 TATGGGCT GAAAGAAAT GCGCATTC CGTCAAGC AACCTATGG
 701 AGTATTCCT GGTCAACCC GACCAATGT CCGTGGCGA ACTGACCA
 751 TACATCGCC ACTTCCAAN NNAAGCCAA ACACCGGA CTGAGCCAT
 801 CGCATGTGG CGCAATGG TTATCCCGC CGACCTGG GTATGGGCG
 851 TCGTGGCTT TCGCTTACC CCGAAMCA CCGGCGCG CATATGGCG
 901 TTAAMATCT TGGCGGCTT CTGTCTGGA TTGCTTGC ACTTGGCG
 951 NCGGCTTC NGATTACCA GCAACTCTA CCGCATCCG CCGTCTGCG
 1001 NCGGCGCT ACCTACCA GCTTGGCT TGTGGCGCT TTGGCTGTA
 1051 CGCAACAGG AAAACGCTA A

This encodes a protein having amino acid sequence <SEQ ID 56>:

1 HMLSRVYIR OHAVNAVYAL LAFALYSFE ELYETGNIG KSGYGVNEMX
 51 GYTLAKMKAR AVELMPLAVL IGLVEXSDL AASSTLYIR ASHSTKEL
 101 LILSOFETI ALATVNLGEV VAPLTSQKAE NIAAALNGK ISTGNTGLM
 151 KENNSINVR EMLPDLTGL IKIWARNDN ELAEVADSA AVNSGSLTT
 201 LKNIRSTLG EDKVEYSIA EXWPISYKR NIMLDVLYRP DQMSVELTT
 251 YIRHLOXXSQ NTRIVAIAM RLYYPRAM UNALVATRT POTTNRHNG
 301 LKRFEGICLG LLEHLNGLF KTSQLYGIP PFLXGALPTI AFALAVLI
 351 RKQERK*

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

orf112a-pep HMLSRVYIR OHAVNAVYAL LAFALYSFE ELYETGNIG KSGYGVNEMX
 orf112-1 HMLSRVYIR OHAVNAVYAL LAFALYSFE ELYETGNIG KSGYGVNEMX
 orf112a-pep AVELMPLAVL IGLVEXSDL AASSTLYIR ASHSTKEL
 orf112-1 AVELMPLAVL IGLVEXSDL AASSTLYIR ASHSTKEL
 orf112a-pep VAPLTSQKAEINIAAINGKISTGNTGLMEKNVINYREMLPDR
 orf112-1 VAPLTSQKAEINIAAINGKISTGNTGLMEKNVINYREMLPDR
 orf112a-pep ELAEVADSAVNSDGSQMLKNIRSTLGEDKVEYSIAEXWPIYSKRNIMLDVLYRP
 orf112-1 ELAEVADSAVNSDGSQMLKNIRSTLGEDKVEYSIAEXWPIYSKRNIMLDVLYRP
 orf112a-pep DQMSVELTTIRHLOXXSQNTRIVAIAMRKYYPRAAMUNALVATRT POTTNRHNG
 orf112a-pep DQMSVELTTIRHLOXXSQNTRIVAIAMRKYYPRAAMUNALVATRT POTTNRHNG

orf112-1 DQMSVELTTIRHLOXXSQNTRIVAIAMRKYYPRAAMUNALVATRT POTTNRHNG
 orf112a-pep LKRFEGICLG LFLNAGRLFXETSQLYIPFLXGALPTIAFLALVLIARLRQERK
 orf112-1 LKRFEGICLG LFLNAGRLFXETSQLYIPFLXGALPTIAFLALVLIARLRQERK

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 13

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 57>

1 ..CGAATAGCG AAGTACGCA CAGGACGGC AAGGTAAAC AGCGAGCAG
 51 TTGGTTCT GTTTCACCTA AACTTACG CGACTTTGC GCGAACTGA
 101 AAGCACTCT TAAACCTTG GTCTGCTCT TGGTTCCCT GAGTATGTA
 151 TTGGTGGCC ATGCCCAAT TACACCGAC AATTCGAC GTAAACCA
 201 GCAGTGCTT ATCCTTAAA CCACACTGG TGGCCCTTG GTGATATTC
 251 AACTCGGA TGGACGGA TTGACACCA ACCGCTA TA CCAATTGAT
 301 GTTACACA AAGGGCAAT GTTAAACCA AGCCATACA ATATCCGT
 351 TGTGTCAAA GCGATTCGC AATGATATT GAAACGATA GCGGTACGG
 401 CTAGCAACT CAGGATATC GTTACCTGG GCGGTCAAA GCGGACGTG
 451 ATATTGCGA ACCCAACGG CATTAACCT AATGACGGC GCTTAAAA
 501 TGTGGTGG GGCATCTTA CTACCGGAG GCGGCAAT GCGAAGGAC
 551 GTGCATGAC AGAATTTAT GTGGCTGAG GCAATGGA CCGTAAAGC
 601 AGCAGGTGG AATGATTAAG GCGTACGAG YTHACCGGG GTACTTGCT
 651 GTCCAGTTG TTTCACGGG AATTTCGAG GTAA..AAT GCGGTTTCT
 701 ACCGCTCTC AGAATGATA TTACCGCAC GCGCAATCA GTCCAGTAC
 751 GCGAGGGGT ACGAACGCA CTATTGCTT TGAATCTGC GCACGTGGC
 801 GTATGTACG CAGACATAC AACTGATTG CCAATGAAA AGCGTACGC
 851 GTCTAA

This corresponds to the amino acid sequence <SEQ ID 58, ORF114>:

1 ..AAVETANSQG KGRAGSSVS VSLTSGDLC GKLTITLKL VCSLVSLSHV
 51 LPAHQLITD KSAFNNQOVY IKTNTGAPL VNIOTPNBNG LSNKRTAFD
 101 VDRKAVIANN DANNRPYVK GSAOLITNEV RGTASRLNGI VTVGGKADV
 151 IIANPNGITV NGRERNVGR GILTGAPDI GRDGLITGD VVARHWTXA
 201 AGNDKGGAX YTGVLAAVA LQGRKXGKL AVSGPOKVD YASGELISGT
 251 AAGTKPTIAL DTMALGMYA DSITLIANKR GVG*

Further work revealed the complete nucleotide sequence <SEQ ID 59>:

1 ATGATTAAG GTTACATCG CATATCTTT AGTAAAGC ACAGACCAT
 51 GGTGACAGA GCCAACTG CCAACGCCA GGGCAAGGT AACAGCCAG
 101 GCAAGTGGT TTCTGTTCA CTGAACCTT CAGGCCACCT TTGGCGAAA
 151 CTGAACCA CCGCTTAAG TTGGTCTGC TTGTTGTTT CCGTGATAT
 201 GGTATGCTT GCGCATGGC AATTAACG CCAATATCA GCACTTAAA
 251 ACCAGAGGT CGTATCTCT AAACCAACA CTGTGCCCC CTGGTGAAT
 301 ATCCAACTC CGAATGAGC CGATTGAGC CACAAACCT ATGACAGTT
 351 TGAATGTCG TCGGGGATC TTATACGCG GTTAATGAG CACGGCTTA
 401 CGTTGTGCT CAAGGCACT GCGCATTTA TTTTACAGA GGTACGGGT
 451 AGCGTAGCA AACTACAGC CATCTGATC GTAGGCGGT AAAGCGCAA
 501 CGTATATTT GCCAACCCA AGCGCATTC GTTAATGAG CACGGCTTA
 551 AAATGTGCG TCGGGGATC TTATACGCG GTTAATGAG CACGGCTTA
 601 GACGTGCG TCGAGGATC TGAATGCT CAGGACATC TATCCGTAG
 651 ACAGAGGT TGAATGAT AAGCGGAGC GCACTACCC GGGGTATTTG
 701 CTGCTGCACT TCGTTCCAG GGAATATAC AGGTAAAA CCGGCGGTT

751 TCTACCGGCT CTCAGAAAGT AGATTACGCG AGCGGGGAAA TCAGTGCAGG
801 TACGGCAGCG GGTACGAAAC GACATCTGCG CTTGTATCT CCGCAGCTGG
851 GGGTNTGTA CCGCCAGAGC ATCAGACTGAG TTGCGATATG AAGAGCGTA
901 TGGCTCAAA CCGCAGCGAG CCGCAAGCAT TGAATTGTGAC
951 TGCTGAGCG CGCATTTGAA ACAGCGGCGG CATGCGCAC ACTCGCGAG
1001 GCACCGAGC TTCACCGACT TATCTCTCA TCGAACACG GMAAAGGA
1051 GGGCAGGCA CATTTATCTC AATGTTGCT GGGATGAGA GMAAGGCTT
1101 ATGTTTATT GAGAGCGAG CAGATATAC TTGCGTAA CAGAGCGCTG
1151 TCGCAATTA CCGCAGCTGC CCGCTACCA CGGATTAATA TCGTGGTAT
1201 ATTTTGGTA TCGAGAGCA AACTATGAT AGCATGCC AAGCGCGCG
1251 TACTCTGTG CCGAGCGGC GTACCGCTG CANGAGGCG AGTATTCAGA
1301 TGGCAGCTA GTTATACAGT TCCAGCAAG GCAACGCC ATTAGGCAAT
1351 ACACAGGTA TTACCGGGC AGATGTTACC GTATTATCCA ACGCACCAT
1401 CAGCAGTTC CCGGTATNG ATGCCAAGA CACGCGAC ATCGAGCAG
1451 GMAACGGT TCTTTGGA GCTTCAACG TACTTCCGA TATCCGCTTA
1501 AAGGAGCA GTATCAAGG CCGCAAGC GTGCTTTAC TGGCAGAGA
1551 CAAATTAAT TCCAAACTA CCAATCTGA TACTCGCG ATCTGTATG
1601 TTCAATAGG TAAAGATCTG TATTTCAATG TTGATTAAGA TTGCTGTGC
1651 GCGACATCC ATTGAAATC GGTATACGT GCCATATTA CCGGCGCTC
1701 TAAACCCCT ACTGCTCTAA AAGACATGG TTGGAGGCA GCTCGCTGA
1751 ATGTACCAA TACCANTCTG GGTACCACT CCGGTATCT GCACATTCAG
1801 GCGACCAAG GCAATATCCA GCTTCCCAAT ACCAASCTA ACGACGCA
1851 GGCCTCGAA ACACCGCTG TCGAGGCA TATGTTTCA GAGGCGCTC
1901 ATGCTGTTT TGCAGAGCGT CATGTATCT TATGBCCA CCGTATGCT
1951 GACTTTACG GTACAGCAT CCGACAGCC AAGCGAGT TCAATGCAAG
2001 ATCGTTGGT AAGGCGCTC TGAACGAGA CAATACCAT ATCATTCAT
2051 CTTGAGNAG TATTAGTGT GTGCGCGCA ACGGTATCA GCTTGGTAC
2101 GGAACACAC GCAATTCAT AAGCGGAAA CACATAGCA TCMAAACA
2151 GCGTGGTAT CCGCTATTA AAMACCTAA GCTGCTAGC AAGCGGCG
2201 CATTGACAT TCAITCGAG CCGGCAITGA GCATAGAA TACACACTG
2251 GAGCTACCC ATNATGCA TCTATGCA CACACGCG GGTATGCT
2301 CACCACTGA ATGCTTACG CACACCGTCA TCTAAGCAT ACACGAGCC
2351 AGTGTATGG AAGACAGAA CCGCTTCTG CCAACAGCT GCGCTATAC
2401 GGTGTATGG CACTAATCG GCGTATTCG CAATATGCG ACNACACCA
2451 GCTGAGGCG GTGCAATCA ACCTTACTG CCGTACGCG CTNCTACAG
2501 GCGGCACAT CAATGGAGT ACGGTTTGA CCAAACTTT GGAAGTAA
2551 GCGATTAAT AACCATTGG CCGACGGCTG ATATTAAG CAGGTAGCG
2601 GCATTAAC ATCGAAGCTG CCAACCGAT CAGTGGCAT ACCGACCTGA
2651 AGCAGAGAA TCGCTGAGG ATTTAGAGG TCTAATAAT ACNCGGTA
2701 GGTGCGCTA GTCTCAATG ATTTAGAGG TCTAATAAT ACNCGGTA
2751 TCTGTTTGA GGAAGACAG ATTTAGAGG TCTAATAAT ACNCGGTA
2801 AAACCTTGT TGTGCGCAC ACCAAGGCA AGTTGATAT CAGACCGTA
2851 ACACACTAT TACCAATTA TTTCTTACA CAAAAGCG CTGACCTAA
2901 CCAAAATCC AAGAATTTG ACACAGAT TCGCGAGTG AAAAAAGCT
2951 CCGTAAAG CAGCTGAT CCAACCTCG AAGAAGAG CAGCGCTC
3001 GCTTCTATA TTCAGGAT CACAGAGGA GTTAAGGTA AARACCTTA
3051 AGCAGAGAA TCGCTGAG CCAAGCTTC CGCAATAT ATGACTGA
3101 TTTCCGACA AGCATGGA ATCAGCGTT TCGATATAC CCGTCTCAA
3151 AAATGAACT TCCAGCGCG ATCAGCGTT TCGATATAC CCGTCTCAA
3201 GCGGCTGCT ATCTGATG AGGCGTAA CAGGATAC GAAATGCGA
3251 ACCGACCTA CAGAGTAC TAGCAAGG CTGCTTGA CAACTGCGA
3301 CTTTGACCG CAGTACAG GGTATGAT CATCGACT CCGCAGCTGA
3351 TGATGACCT ATTTATATG GTGCTTCCA AATCAAGT CCGTACGCA
3401 GCATAGCAT CAAGCGCAT AGTATNTG TACTGAGCG TGCACAAAC
3451 GATGCTCTA CTTCTTAA ACCAAGGT AAGAGGGA AACTCATAC
3501 AAAAAAGG TTTACAGCA CCGCGAGCA CCGTATAT CAGGCAAGT
3551 TCGAGTAC CCGCAAGCG ATACGCTTC AGCGAGCG CAACTGGA
3601 GCTATACCA CCGCTTCA TGCCTTGA GGTAAAGTA CCGTGTGCT
3651 GGTGTAGAG CTGCACTG TGCAGAGA AGGCTCAC AAGCAGAT
3701 TGGATGCTA AAAAGCGCG CCGTTATTC GCAATCAAGT AGGCAAGT
3751 ATTAACATA AAGACGAT CAGCAAGAC AATTTSCGT TCCGCTGCT
3801 CCGCAACT GAGCGACCG GTTACAGCTG GATACCGT CTCAGAGTA
3851 CCGAATTA ACACACCTG CCGCTGCG CACTTACCG AGGTGAGCG
3901 GAAAAGGCC GTGCGGATG GAAATATAT CTCAAAGCA TTGTGAACG

5

10

15

20

25

30

35

40

45

50

55

60

3951 TATCCAGTGG GAAGAAAA TAGAAACCAA CTCACCGTA TGGCAGAAC
4001 AGGCGGAGC CCGCAGCACT ATCGAAACG TGAACCTGC CAGCTTCGA
4051 AGCCCTACT CCGCCTAAT GACCCGCCG GTGGCTATA TGTGACAT
4101 TCGAAGCG ATTITGAHA CCGAATCGA AAGCTGGCC AARCGCCG
4151 AGTATGCTTA TCGCTTACGA TAAATGGAC TATAAGCAG AGGCTTAC
4201 CAGGTCAAC TGTCAAGTAG CCAAAAGCT CAATCGAAC
4251 CAGAGCGGT CGAGGATG TTACATAT CTGACCGCA CTGACTATG
4301 GATAGGCGC ACCCGAGCG GCGGTGTAG CCGCTTACG AAGTAGTAC
4351 CCGCAGCTG CCGAAGCAG CCGCACACG ACAGCAGCAG CTACTACGT
4401 TTCTACAGC ACTGCCATC AARCGCTG TTAGCTCC TTGTATAGC
4451 AAGCAGCT ATCATCATC AATATNAG GTAGTCTG CAAAGCTTG
4501 AAGATCTG GCACAGTG TACGTNAG CAGATTGTA CTTCGCTG
4551 GAGGCGGT GCATTAATC ATGATGGCG AGATTTGCC CANTTGA
4601 GCAAGGTAG AACCAATC TTACAGTA CGGCAATCA ACTATTGCG
4651 AACCTGAG GAGACTGCG TACCAATC AGTATGAG GTATCTAGC
4701 TGTATCAT AGGCGCTCA CCGCGGAG CCGTAAGAG AACTTAGCA
4751 ATGCGCAT AGGAGATG GTTAAAGT TCCNAGAGA AGCGGCGAG
4801 AATATCAA CAACTTACG CCGATTAT GTGCCAAC AGTTGCGCA
4851 CCGTTGCT GGTGTGTA GCGATTGT ACAGGAAA GTTAAGAGC
4901 GCGAATCG CCGCAGCT GCGAATCG GCGAATCG TATGCTGCG
4951 GCGAAGC CTGCTACT CAGCATCG GAAAGCATA AGGTTATAG
5001 TTACTGAG ATATTGCG CAGGCTGCG GCACTCAAC GCGCGCATG
5051 TGAATCTG GCGAATCG GCTAGTGG CGTACTGA TAATGCTTG
5101 TAATTTGCA GTACCTTAC CANTCGAA MAGCATCAG CCGAAGC
5151 CCGAAGC GCATGGA AATATCAA AGGATTAT CCGCATAG
5201 CAGCAGTGC GATGCTAT CCGAGGATA AGGATGCT CATTGATA
5251 AGCATATC GTATGCTAT CCGAGCGG ATGTGATA CCGATAGG
5301 GGTATCTG CAGGTTGA CAGCTCGCT GATGCTACA GCGGTAAT
5351 TAGTATCAG CACTGCTAT GCTAATCTT TGTGTTGAC TGTGCTG
5401 ACTCAGCTG CCGAAGCGG CCGGAGATC CCGAAGCGT CCGTACGCT
5451 AGCAGCTG TGGAGCGC CTGTGGCG GTTGTGAA CGAAGCGG
5501 CCAAGCAG TATACACG CAGACTGA MAGACTGA TGGTACTA
5551 CAGATCAA AATATAGG TCGTAAAT ACAGATTA ATATAGCA
5601 TAGTACTCT CGATATAC CAATGAGCA AAGGACAA CCGTATCTG
5651 CTGCTTTGA CGATTTCT GAGGCGCAT TCCATAGCC TATGCGAT
5701 AAGCTTAC TTTTACAT CTCCCAAT CANTGAGG TTATCTCA
5751 AATATNAG GTGTTTCT CCGCTATC GATGCTCT TATGCGAT
5801 ATATGCGAG GTAGTATGA GAAAGTGA TTGCTACT TGTCTTAA
5851 AAGGTGAC AACCAACAC TACATTA GATTATAC ATAGTCAG
5901 AATTTGAT ACTACATAC CAGTAAGG AACTTA

This corresponds to the amino acid sequence <SEQ ID 60; ORF14-1>:

1 MNKGLHRIIF SKKSTHVAV AETANSQKQ KQAGSSVSVS LKTSGLCGK
51 LKTLTLVLC SLVSLHVLV AHAQITDKS APKNOQVVL KNTGAPLVN
101 IOTPNRGLS HNRVTOFDV NKGAVLNDR NNPFWKGS AOLINERVG
151 TASLKLNGIV VGGQADVII ANPGITVNG GGFNVRGI LITGAPQIG
201 DGALTGVDR QGTLTGAAG WNDGGADYT GLAVALAQ KRLQKNAV
251 STGPQKDYA SCEISATGA GTPKTIADT AALGHWYADS ILLANEKV
301 GVRNAGTLEA AKQLIVSSG RIENSRIAT TADGTEASPT YLSIETKRG
351 AAGTFISNG RIESKGLVI ETGEDISLRN GAVQNGSR PATTVLNAGH
401 NLVIESKINV NNAKGPATLS ADGRVLIKEA SIOTGTVYS SSKGNALGN
451 NTRITGADVT VLSNGTSSS AVIDAKOTAH IEAGRPISLE ASVTSDIRL
501 NGGSIKGGK LALLADNIT AKTNLNTGK NLYVHTKGL RNSGNLHIQ
551 ASHLKSONA AHITGSKTL TASKMGVEA GSLVNTNL RNSGNLHIQ
601 AAKGTQLRN TKLNAKALE TTAOGHIVS DGLHVSADG HVLNANGA
651 DFTGHTLTA KADYNAGSV KRLKADNTN ITSSGDIIL VAGNQLQGD
701 GRQNSNGK HISIKNGVN ADKLNVAH KSGALNHSO RALSIENTKL
751 ESTHNLHNA QHVERTLQV DAYARHLSI TGSOIWONDK LPSANKLVN
801 GVALNARYS QIADNTLLRA GAINITAGTA LYRGNINWS TVSTKLTEDN
851 BELKPLABL NIEAGSGTIT IEPANRISAH TDSIKTGGK ILLSAKNGA
901 GAFSAVSLL EAKGNIRLYT GETDIRGSKI TAGKRLVLI TKGRLNIEAV
951 NNSFNYFT QKALNQRS KELEQIOL KSSPKSKLI PTLQERDL
1001 AFYQIAINKE VNGKPKGKE YLOAKLSNQ IDLSAOGIE TSGSDITASK

45

50

55

60

1051 KILNHAQVL PKADSEAA ILIDCITDQ EICKPTYSK YKAAALNPS
1101 BLTGRTGSI HAAALDDAR IIGASEIKA PSSGIDIAH SDIYLAQON
1151 DAITFLTKG KSGKILKTRK FISTRDHLH PAPBELTAN IYLAQAGNE
1201 ANTRENAQA KGYVLVAGEE LQLEAGIEH KHELDYQSR RATICVAGS
1251 NYSKELNET KLPYRVVAGT AATRSQMDVY LGEFEKTL ACADIDAVG
1301 EKARADAKI LKGIWRIS EKEKTNSTV WQRQARGST IETIKPSFE
1351 SPTEPLTAP GGIYIVDIPAG NLKTEIELKA KOPEYALIKO LQYVAGST
1401 OVOLAYDKD YKQGLIRBAG ALATVILYTA LYTGATATA GQVAGSST
1451 AAAAGTAAT TAAATVSTP TAMOTAAAS LYQANASTI NKQDYGAL
1501 KDGTSQDYK QIVTSALITG ALNQGDVIA QUNSVARTEL FSTIQNTIA
1551 MUGRLATNL SNAIGSAGIN TAVNGSLKD NLGNAALGAL VSTQGEAAS
1601 KIKTESDQY VAKQFAHALA GCVSGLYQK KNGALGAY GEIYASHLG
1651 GNPATLSDA EKHVATYSK IYAGVAAAN GGVYTAAMA AEVAVNNA
1701 NEDSTPAK KHQQRKPKT ALEKIQGN PAHAGANTY PODKALMI
1751 SNIRNGITG IYITSYVGA AGWAPLIGT AKLAISTCH ANPSCGTAT
1801 TQAAEGAGI ATGAVTVGNA WAPYVGLAK AKAAQALPT QYKELDEL
1851 QESKNGAVN TRINIANST KITPRQTOQ PVASQFEVY EGHNRPYAN
1901 MRSVFTJSE ELKVILQSKN VVSPVSMTP DQGYHRTVDY KVLGTTSIK
1951 EGGOPTTIK VETDKSGLI TYYPKVN*

Computer analysis of this amino acid sequence predicts a transmembrane region and also gives the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF114 shows 91.9% identity over a 284aa overlap with an ORF (ORF114a) from strain A of *N. meningitidis*:

25	orf114.pep	10	20	30	40
	orf114a	10	20	30	40
30	orf114.pep	10	20	30	40
	orf114a	10	20	30	40
35	orf114.pep	10	20	30	40
	orf114a	10	20	30	40
45	orf114.pep	10	20	30	40
	orf114a	10	20	30	40
50	orf114.pep	10	20	30	40
	orf114a	10	20	30	40
55	orf114.pep	10	20	30	40
	orf114a	10	20	30	40

The complete length ORF114a nucleotide sequence <SEQ ID 61> is:

orf114a GVNAGTLEAHQILVSSGRISNSGRATTAATDGTASPTYLKLETFEGKAGTFSNGG
310 320 330 340 350 360

1 ATGATAAAG GTTTACATCG CATATCTTT AGTAAAGC ACAGCCACAT
51 GGTTCGACTA GCCGAAATG CCAGACAGCA GAGCAAGGT AACACAGGAG
101 GCGATTCGCT TTCTGTTCA CTGAACATCT CAGGACACT TTGCGGCAAA
151 CTCMAACCA CCTTAAC CTGTGCTCG TCTTGTTGT CCGTGCAAT
201 GAAATTCNUN NNCNNNCC AAATTACAC GCACATCA CTTGGCTTAA
251 ACCMAAGCT GGTATCTT AAACACACA CTGGGCCCC CTGCTGAA
301 ATCCAACTG CGATGAGG CGGATTGAC CAAACCGCT ATACGACAT
351 TGAITGAC ACAAAGG CAGGTITAA CAACACCGT AACANATC
401 GTTTCTGT CAAGGCGT GCGCATATG TTTGACAGA GATGACCGT
451 ACAGGTACA AACTCAAGG CATGTTACG GTNAGCCGCT AAAGGCCA
501 CGTATTAT GCACACCCA ACAGCATTC CTTATGCG GGGGCTTTA
551 AAATGTGG TCGGGCTC TTAACTATG GTGCGCCCA AATCGGCAA
601 GACGTGAC TGACAGAT TTATGTGCT CAGGACACT TGACGTAG
651 AGCAGAGGT TGGATATGTA AAGCGGAGC CGACTACAC GCGTACTTG
701 CTGTGACGT TGTCTTGGG GGGAAATTC AGGTTAANA CTGCGGCT
751 TCTACCGCT CTCAAAAT AGATTACGC ACAGGCAAA TCAGTGACG
801 TACGACAG GATGAAAC GCACTATTC CCTGATAC CCGGCACTG
851 GCGGTATGTA CCGCCACAC ATCACATGA TTGCAATGA AAAAGGCTA
901 GCGGTCAAAA ATGCGGGAC ACTGACAG GCGACATAT TGAITGAC
951 TTGTGTAGC CGCATTTGAA ACAGGCGCG CATGCGCAC ACTGCGCAG
1001 GCACGAGC TTACCCACT TATCTNCA TCGAACAC GGAANAAGA
1051 GCMNAGCA CATTTATC CATGTGCT GCGATGARA GMAAGCTT
1101 ATGTGATAT GAGACGGGAG AAGTATCAN CTTGCGAAC GAGACCTTG
1151 TCGAATATA CGCGACTGC CAGCTACCA AACTATGTA AGGTGCTAT
1201 AATTGTGTA TTGAGATGA AACTATGTA AACTATGTA AGGTGCTAT
1251 TAATCTGC GCGCGGCTC GTACTACAT CAGTATGTA AGGTGCTAT
1301 CGGCGACTGC GGTGACAC TCGACCAAG GCGATACGA ATGTGCTAT
1351 AATACCGCTA TTATGCTTA AAGGTACAC GATATACGA ATGTGCTAT
1401 TGGACGCTG GCTGATGTA AGGTGACCA CAGTATGTA ATGTGCTAT
1451 GCAACCGCT TTCTTATGA ACCTGACCG TTGCTTCCA CATCGCTTG
1501 ACAAAGGCTA AACTAAGG CGGAAGCAG CTTCTTTA TGGACGCA
1551 TAACATTA CTGCAACCTA CCAATGTA TAATCTGCG CATGTATG
1601 TCAATACAG TAATATCTA AATTGATG TAATCTGCG CATGTATG
1651 GCGACATC ATTTGAAT GGAATACCT TGTATGTA ATGTGCTAT
1701 TAAACCTC ACTGCTCA AAGATGCG TTGTAATGA CCGGACAG
1751 ATGTACCA TACATCTG CTTACACT CCGTATCT TATGCTAT
1801 GAGGCAAG GCAATTTA GCTTGCAT ACACATCTA AGGACGCA
1851 GGTCTCGA ACCACCGT TCGAGGCA TACGTTCA GAGGCTC
1901 ATGTGCTG TCGACGCT CATGATCT TATGCTAT CCGTATCT
1951 GACTTACG GTACATCT CTTGACAC AAGGCGAT TCMATGAG
2001 ATCGTTGT AAGGCGCT TGAAGCAG CAATACCAT ATCATCTAT
2051 CTTCAGAG TATATCTG GTTGGCANN TCGATATCA GTTGTGCT
2101 GGAACGAC GCAATCAT CAAAGCAA CAGTATCA TAAACAA
2151 CGGTGTAT GCGCATTA AAACCTTA GCTCATGCA TAAACAA
2201 CATTAACAT TCAATCTG CCGGATAT GCAATGAA TCAATGCT
2251 GAGTCAAC ATAAATGCA TCTATGCA CAAACGAG GGTATGCT
2301 CAACAGTA GATCTATG CACACCTA TCAACATG ATCATCTAT
2351 AGATTGGA AAGACAAA CTGCTCTG CAACAGCT GTTGTGCT
2401 GGTATATG CACTATCT GCGCTTTC CAATTCGG ACACACAC
2451 GCTGAGAG GGTGATTA ACTTATCT CCGTACCC CTATGACAC
2501 GCGGCAAT CATTTGAT ACCGTTGCA CCAAGCTT GAGTATAT
2551 GCGAATTA AACATGCT CCGAGGCTG AATATGAG GAGTATAT
2601 CACATTAAC ATCAACCTG CCAACCGCT CATGCGCT ACAGCTAT
2651 GCATCAAC AGCGGAAA TTGCTGTG CTCAAAAG AGAATATG
2701 GTGCGCTTA GTGCTAAT TTCTCATG ATTATAGG TTCTAAT
2751 TCTGTATGA GGAAGACG ATTATAGG TTCTAAT ACAGCTAT
2801 AAACCTGT TGTGCTAT TTTTATCA CAANAAGN NNNGCTCA
2851 AACACTAT TACATATTA TTTTATCA CAANAAGN NNNGCTCA
2901 CCANAATCT AAGATATG ACAGCAT TGGCAGTG AAAAATCT

2951 CGGTAATACG CAGCTGATT CCAACCTGCG AAGAGBAGC CGACCTGCTC
 3001 GCTTTCTATA TTACGCCAT CACAGAGAA GTTHAGGTA AAAACCCAA
 3051 AGCAAGAA TACCTCGAG CCAAGCTTC GACAAAT ATGACTTCA
 3101 TTTCGCGCA AGGATCGAA TTTCAGCGT CCGATTATAC CGCTTCCAA
 3151 AAATCGAACC TTACAGCCGC AGGCTATTG CCAAGAGCAG CAGATTGCA
 3201 GCGGCTGCT ATTCTGATT AGGCTATAC CAGCATAT GAAATGGCA
 3251 AGCCCACTA CAGAGTCTAC TACGACAAAG CTGCTCTGAA CAGCTTCA
 3301 TTGTCACCG CAGAGTACGG GGTAAAGTAT CATCGCTGCA CGGCATCTGA
 3351 TGATACAGT ATATTATTCG GTGATCTCGA ATCAAAGCT CCTCAGGCA
 3401 CGATACAT CAAGACCAT AGTGATATTG TACGGAGCG TGCACAAAC
 3451 GATGCTATA CTCTTTANA AACCAAGT AAAAGCGCA NAATATACG
 3501 AAAACAAAG TTATACAGCA CCGCGANCA CCGTATTATG CAGCGCCNG
 3551 TTGACCTGAG GCGCAACGCT ATCAGCTTC AGGAGCGCG CAACATCGAA
 3601 GGTATACCA CCGCTTCAA TGCCCTTGA GGTAAAGTA CCTGCTGTC
 3651 GGTGAAGAG NTGCACTGCG TGCAGAGA AGGATCTCAC AGGCACAGT
 3701 TGATGTGCA AAAAGCGCG CCGTTATCG GCATCAAGT AGGTAGAGC
 3751 AATTACAGTA AAAACGACCT GAACGAACC AATTTGCTG TCCGGCTGCT
 3801 GCGCAANT GAAGCAGCC GTTCAGGCTG GGTACCTG CTGCAAGTA
 3851 CGATATCAA AACACGCTG CCGGTGCGC ACATTCAGC AGGTGANGC
 3901 GAAAGGCC GTGCTGATCG GAAATATTC CTCAGGCA TTGTGAACCG
 3951 TATCAGCTCG GAGAAACAT TAGAAACAA CTCACCGTA TGTGAACCG
 4001 AGCGCGGCG GCGCAGACT ATCGAAGCG TAAACTGCC CAGCTTCAA
 4051 AGCCCTACT GCGCCTAAT GTCCGACCC GCGGATATA TCGTGACAT
 4101 TCGAAGGCC AATCTGAAA CCGAATATC AACTCTGCC AACAGCCCG
 4151 ATGTGCTCTA TCTGAACAG CTCCAAGTAG CGAAGACAT CAATCGAAT
 4201 CAGGTGCGC TTGCTTACGA CAGATGGGAC TACAAACG AGGCTTAC
 4251 CGAAGAGT GCGCGATTA TCGACTGCG CGTACCTG GTCACTGAC
 4301 GCGAGAAC CCGAGCGCTA TTGGGATTA AGCTGCGC CCGCGCGCA
 4351 ACCGATGAG CATTCGCTC TTGCGCAGC CAGCTTCCG TATGCTTAT
 4401 CACAAACAA GCGATGCTCG GCAAACTCT GAAGAGCTG GGCAGAGCA
 4451 GCACGGTGA AATCTGCTG GTTCGCGCG CTACGCGAG CGTAGCCGAC
 4501 AATTCGGCG CTTCGCTACT GANCAATG AGCATAAG AGTGAATCA
 4551 CAACCTGAC GTAACCTAG CCAATGCGG GCACTGCGC ACTGATLaa

This encodes a protein having amino acid sequence <SEQ ID 62>:

1 MNKGLHRIIF SKRSTWAV AETNSOGK KQAGSSVS LKTSGLGCK
 51 LKTLTKLVC SLVSLSHXXX XXXQITDKS APKXQVIL KNTGAPLVN
 101 IOTPNRGAS NHRVQDFVD NKGAVLNDNR NNHFLVGS AOLILNEVRG
 151 TASKUNGIVT VGGQADVII ANPGITVNG GGFNVGRGI LTIGAPQIGK
 201 DGLTGFDRV QGLTVGAG WNDKGGADT GVLARVALQ GKLOKRLAV
 251 STGPQKYDVA SGEISGATA GTREPTALDT AALGSHVADS ILLIAEXKGV
 301 GYKNAGTLEA AKQILVTSAG RIENSGRIAT TADGTEASPT YLXIETKGG
 351 AXGTFSNGR RIESKGLLVI ETGEDIXLRN GAVYQNGSR PATTYLAAGH
 401 NLVIESKTVW NNAKGSXNLS AGRTTINDA TIOAGSSVYS STNGUTXGE
 451 NRIITAEVNT VLSNGSIGSA AVIEAKDTHA IESKPLSLSE TSTVASNRL
 501 NNGNTKGGQ LALLADNIT ANTIENLTPG NLYVHTKOL NLNVKDKLSA
 551 ASHLKSDNA AHITGTSKTL TASKONGVEA GLAVNTNL RTNSGHLHTQ
 601 AKNGIQLRN TKUNAARALE ITALQGNIVS DGLHVSADG HVSLLANGNA
 651 DTGHTMLTA KADYVAGSVG GKRLKADNTN ITSSSGDITL VAXXQIQGLD
 701 GKORNSINGK HSIKNGNGN ADKLNINVA KSGALNHS RALSTENTKL
 751 ESTHNTLNA OHEVTLNQV DAYAHRHLSI XGSIQWONK LFSANKLVAN
 801 GVLXNARKS QIADNTLNA GAINTLTA LVRKGNINS TVSTKLEON
 851 RLTGTGVSII HAAALDAAA ILIDGITDQY EIGKPYKSH YOKAALNKPS
 901 GAXSAQVSSL EAKGNILVLT GXTDLRGSKI TAGNLVAT TKGLNIEAV
 951 NNSFNVFT QYXXILNKS VELEQIQL KKSXSKSLI PLQEDRAL
 1001 AYTQALNKE VGRKPKGKE YLQKUSQIN IDLSAOGIE ISGSDITASK
 1051 KUNLHAAGVL PAAASDAAA ILIDGITDQY EIGKPYKSH YOKAALNKPS
 1101 RLTGTGVSII HAAALDAAA ILIDGITDQY EIGKPYKSH YOKAALNKPS
 1151 DAYFTLTKG KSGXIRKTK FTSTYXHLIM PAVELTANG IFLQANGTE
 1201 ANTRFNAPA KTYLVAGGX QOLLAEEGIH KHELDVOKSR RFIGIKVGS
 1251 NYSKNEUET KPELVAVQX AATRSQDWT LEGTEFKTL AGADIQGVX
 1301 EKARDAKII LKGVINRIQS BEKLETNSTV WOKQAGST IETLKLPSFE
 1351 SPTPKLSAP GGVIVDPKG NLKTEIKLS KQPEVAYLQ LQVAKNNIN

1401 QVQLAYDRD YKQGLTEAG AAILALVTV VTSAGTGAAG LGINGAXAAA
 1451 TDAAFASLAS QASVSFINNK GVGKTKLKL GRSSTVKNLV VAAATAGVAD
 1501 KIGASALXNV SDQKPINLIT VLNKNGQCR TD*

ORF114-1 and ORF114a show 89.8% identity in 1564 aa overlap

5 orf114a.pep MNKGLHRIIFSKRSTWAVAVETANSOGKQAGSSVSLSKTSGLGCKLTKTLKLV
 orf114-1 MNKGLHRIIFSKRSTWAVAVETANSOGKQAGSSVSLSKTSGLGCKLTKTLKLV
 10 orf114a.pep SLVSLSHXXXQIITDKSAPKXQVILKNTGAPLVNQTTPNGRGLSHNRVYTFQD
 orf114-1 SLVSLSHVLPNHAQIITDKSAPKXQVILKNTGAPLVNQTTPNGRGLSHNRVYTFQD
 15 orf114a.pep NKGAVLNDNRNNPFLVKSQAOLILNEVRGTASKLNGIVTVGGQKADVLIANP
 orf114-1 NKGAVLNDNRNNPFLVKSQAOLILNEVRGTASKLNGIVTVGGQKADVLIANP
 20 orf114a.pep GGFNVGRGIITIGAPQIGKDGALTGDFVRQGLTVGAAGWNDKGGADTGVLARVALQ
 orf114-1 GGFNVGRGIITIGAPQIGKDGALTGDFVRQGLTVGAAGWNDKGGADTGVLARVALQ
 25 orf114a.pep GKLOKRLAVSTGPQKYDVA SGEISAGTAAKTPTIALDTAALGSHVADSLTIAEXKGV
 orf114-1 GKLOKRLAVSTGPQKYDVA SGEISAGTAAKTPTIALDTAALGSHVADSLTIAEXKGV
 30 orf114a.pep GYKNAGTLEAAKQILVTSAGRIENSGRIATADGTEASPTYLXIETKAGXTFSNGG
 orf114-1 GYKNAGTLEAAKQILVTSAGRIENSGRIATADGTEASPTYLXIETKAGXTFSNGG
 35 orf114a.pep RIESKGLLVIETGEDIXLRN GAVYQNGSRPATTYLAAGHNLVIESKTVNNAKGSXNLS
 orf114-1 RIESKGLLVIETGEDIXLRN GAVYQNGSRPATTYLAAGHNLVIESKTVNNAKGSXNLS
 40 orf114a.pep AGRTTINDATIQAQSSVYSTGDTXGENTRIENVTLSNGSIGSAVIEAKDTHA
 orf114-1 AGRTTINDATIQAQSSVYSTGDTXGENTRIENVTLSNGSIGSAVIEAKDTHA
 45 orf114a.pep IESKPLSLSESTVASNIRLNNGNIRKGGQKALLADNITAKTNTINPGLNVHTKOL
 orf114-1 IESKPLSLSESTVASNIRLNNGNIRKGGQKALLADNITAKTNTINPGLNVHTKOL
 50 orf114a.pep NLNVKDKLSAASISHLKSDNAHITGTSKTLTASKONGVEAGLLNVTNLTNSGHLRQ
 orf114-1 NLNVKDKLSAASISHLKSDNAHITGTSKTLTASKONGVEAGLLNVTNLTNSGHLRQ
 55 orf114a.pep AAKNGIQLRN TKUNAARALETTA LQGNIVSDGLHVSADGHSVLLANGNADEFTGHTLTA
 orf114-1 AAKNGIQLRN TKUNAARALETTA LQGNIVSDGLHVSADGHSVLLANGNADEFTGHTLTA
 60 orf114a.pep KADYVAGSVGKRLKADNTNITSSSGDITLVAXXQIQGLDQKORNSINGKHSIKNNGN
 orf114-1 KADYVAGSVGKRLKADNTNITSSSGDITLVAXXQIQGLDQKORNSINGKHSIKNNGN
 65 orf114a.pep ADKLNINVAHKSALNTHSDRALSENTKLESTHNTLNHAQHERVTLNQVDYAHRLSI
 orf114-1 ADKLNINVAHKSALNTHSDRALSENTKLESTHNTLNHAQHERVTLNQVDYAHRLSI
 70 orf114a.pep XGSIQWONKLPKSNKLVANGVLAAXNARYSOIADNTLTLRAGANLITAGTALVKRGNINS
 orf114-1 XGSIQWONKLPKSNKLVANGVLAAXNARYSOIADNTLTLRAGANLITAGTALVKRGNINS
 75 orf114a.pep TVSTKTLTDNAELKPLAGRLNTEAGSGTLLTEPANRISAHTDLSITGCKULLSANGNA
 orf114-1 TVSTKTLTDNAELKPLAGRLNTEAGSGTLLTEPANRISAHTDLSITGCKULLSANGNA

5
+ A + + + A SG+LHI +A +Q NT L N + A+E++
Sbjct: 563 IARERLDIGAREIENREALLSSGDHLIGSALNGSRVQVAGTANSHRSAAIES--- 619
Query: 626 GNI 628
Sbjct: 620 GNI 622

10
Score = 37.5 bits (85), Expect = 0.53
Identities = 87/432 (20%), Positives = 159/432 (36%), Gaps = 62/432 (14%)
Query: 239 LOGKLGKRLAVSTGPKVDYASGEISAGTAGTKHTIALDTAALGGHYADSTILIAEXK 298
LOG LOGKRL+ + G + +G I A K A + + S T +
Sbjct: 1023 LOGDLOGKRIFAAGSDITN--TSGIAGNALLK-----ASNIESRSETSRQNE 1072

15
Query: 299 GVGKNAGTLEAARQLIVTSGR1--ENSGRIATTDGTEASPTYLXIETEKAGX-TF 355
V+N G + A L +G + + I TA E T + G T
Sbjct: 1073 OGVRNIGRV-AGVILTRONGSVLLDAGNNIVITIS-----ELTNQSEGGTV 1120

20
Query: 356 ISNGRIEKGILLVETGEDIILNRGAVQNNGRSRATTVLNAGHNLVIESK-----T 408
++GG I S + I + V++ + +T+ G NL + +K
Sbjct: 1121 LNAGGDIRSDTIGSRNONTIFDSNVYIRKEQNEVGTITRIG-NLSLWAGDIRIRAA 1179

25
Query: 409 NVNNAKSNLSAGGRITINDATIQASS-----VYSTKGDTXILGENTRIIAENV 460
V + +G L+AG D ++AG + Y+ G + TR +
Sbjct: 1180 EVGSEGRKLKAG-----ROINVEAGKHTETEDLKYRSGGGIQQWTRHLKQNG 1234

30
Query: 461 VLSNGSIGSAVAIEARHTAHIESGRPLSETVSNIRLNNNGIKGKOLALLADONIT 520
+G++ +I +G + + T+ S NN +K + + A+ N
Sbjct: 1235 QAVSGTLOGKEILLVSGRDITVGTSGNIADNHTLS--KNNIVLKAETRSRSEHNKK 1292

35
Query: 521 AKTTNLMPG-NLYVHTKCOLNLYNDKLSAASIRHNSDN-----AAHITGTSKLTA 572
K+ + + G + KD N + +S + S N H T T T++
Sbjct: 1293 EKSLGMSGGGIGTAGSKKQTNRSETVHTSVESVGLNGTILISACKHYTQTGSTISS 1352

40
Query: 573 SK-DHGVGAGXXXXXXXSGNLHIQAKG-----NIQLRNLKNAARALETTALOG 626
+ D+G+ +G + + KG ++ + NT + A A++ G
Sbjct: 1353 PGDVGISGSGKSIDAQNRYSDQSKQVYEQKVTVAISVPVYVTVMGADVAKVQTGV 1412

45
Query: 627 NVVSDGLHAYSA 638
+ ++A++A
Sbjct: 1413 KSKNSRVNMAA 1424

Amino acids 1-1423 of ORF114-1 were cloned in the pGex vector and expressed in *E.coli*, as described above. GST-fusion expression was visible using SDS-PAGE, and Figure 5 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF114-1.

Based on these results, including the homology with the putative secreted protein of *N.meningitidis* and on the presence of a transmembrane domain, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 14

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 63>

5
..CGCTTCATTC ATGATCAAGC AGTCGCGAGC AACATCGGGC GCGGCAAAAT
51 CTGATATGCA GCGGGCAGG ATATCAATGT AGCGGCGAAG AGCTTATTT
101 CTGATAGGG CATTGTTTAA AAGCAGATAC AGCAGATCGA TATTTCTACT
151 GCCCATATC GCTATACCGG CAATGATATC CAGAGAGCA WAAATCTAGG
201 CGTCATGGGT ACTGGCGGAT TGGGCTTTAC TATCGGTAAC CGGAARACTA
251 CCGATGACAC TGATCGTATC ATATTGTSC ATACAGGACG CATTATAGGC
301 AGCTGAAATG GAGACACCGT TACAGTTGCA GGAACCGCT ACCGCAAAC
351 CGCAGTACC GTCTCCAGC CCGAGGGGGC CANTACGCTC ACAGCCAAW
401 GCATAGATGT AGAGTTGCGA AACAAACGGT ATGCCACTGA CTAGCGCCAT
451 ACCGAGGAA CAAAAGGCC TTACCGTGC CTTCAATGTC CCGGTTGTC
501 AAGCTGCAAA AACTTCTATA CAAGCAGCCC AAATGTGGG CAAAGTAA
551 AATAACCGG TTRATGCCAT GCGTGCAGCC AATGCTGCAT GGCAGATTA
601 TACAGCAAA CACAATATGC AACATTTGC TCAGAGCAGC AGTGGCGGAC
651 AAGGTCAAA CTAATATCAA AGCCCCNGTA TCAATGTGTC CATTAC.TAC
701 GCGCAACAGA AAGTCGTAA CGAGCAAAA AGACATTACA CCGAGCGGC
751 AGCAAGTCAA ATATATGCGA AAGGCAAA CACACTTGGC GCACAGAA
801 GTGGGAGCA GTCAATATATC AATATTACAG GTTCGATGT CATCGGCCA
851 GCAGGTACTC C.TCTATTCG CGACAACCA ATCAGACTCC AATCTGCCA
901 ACAGAGCGCG AGCGACAAA CCGCAAAAC TACCATCCGA AGCGCGGGG
951 GCGTACGTG CAAATATAGC AACGCAATCA GGTITGGAAT TACCGCGGA
1001 GGAATATCG GTAAGGTAA AGACAGAGG GGAAGTACTA CCGCAGGCA
1051 ACAGAGCGCG AGCGACAAA CCGCAAAAC TACCATCCGA AGCGCGGGG
1101 GATCCACCC TCAAAGTGT GCAGCTCATC GCAAAAGCA TACAGGAGG
1151 TAGCGCAAC CTGCATATAG AAGTGTCCA AGCATATGAA ACCTATCAG
1201 CCAACAGCA AAGCGCAAT GTCCAGTTE ACTGTGCTGT ACGGATCTCG
1251 TACAGCGCG AGTATACGCC AAGCAAACT CAAGCAGAC GACTGCTCCG
1301 TGACCGGCA AAGCGGTATT TATCGCGAG AGACGGGCTA TCAATATAA
1351 GTVAGAGCA ACACAGCTC YAGGGCGT ATCATCAGT CTAGCAGAA
1401 CGCAGATAT AAGGCAAAA ACTTTTTC GAGCGGCACC CTACTTGCCA
1451 GACAGATTCA AAACCAACG CGCTAGAGAG GCAGAGCTT CCGCATAGCC
1501 AGGAGCTTCG ACTTGAAGC GCGTGGGAG GGCAGGCTTA CCGCAAAAC
1551 AGCAGCGCT ACCGACAGA TAAGCGCGG GCGCGCTAC GCGAGGAGC
1601 GAGACAGCA AAACAGACC ACCCGAGCG GGTCAACAC CCAACATA
1651 CACATACCG AGAAGCGCG ACAAATTGCG CAGACAGCA GACTGCGAA
1701 AGAACCGNA GCGGTATCT ACACCGCAT CGACACCGAA ACTCGGATC
1751 AACACTCAGG CCATCTGAAA AAGAGCTTCG AC...

This corresponds to the amino acid sequence <SEQ ID 64; ORF116>:

1
..RFTHDAVGS NIGGRHIVA AGQDINVRGX SLISDKGIVL KAGHDIST
51 AHRHYTNEY HESXSVYMG TGLGFTIGN RTKTDIDTDT NIVHTYSIIG
101 SLNGDVTVA GRYROTGST VSSPEGRNV TAKXIDVEFA NRYATDYAH
151 TQEQKGLVA LNPVPTQAAQ NFIOAQNUG KSNRNVNAH AANAAMOSY
201 QATQHQOQFA PSSAGQGN YNQSPSISV IYGEQKSN EOKRHYTEAA
251 ASQILGKGT TLAATGSGEQ SNINITGSDV IGHAGTXLIA DNHRLQSAK
301 ODGSEKSNK SSGWNAVYRX KIGNITRFGI TAGHIGKX EGSGTTHRH
351 THVSTGKTT TIRSGGDTTL KGVQLIGKI QADTNLHIE SVQDTYFQS
401 RQNDLKGGI ITSSQSAEDK GRHLFQTATL TASDIQWHSR YEGRSFTGG
451 RQNDLKGGI ITSSQSAEDK GRHLFQTATL TASDIQWHSR YEGRSFTGG
501 SFLANGWDG TVTDKQGRPT DRISPAAGYG SDGSKNST RSGVNTNHI
551 ITDEAGQLAR TGTAKETA RIYTGIDTET ADQSHGLKN SFD...

Computer analysis of this amino acid sequence gave the following results:

Homology with *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF116 and *pspA* protein show 38% aa identity in 502aa overlap:

Orf116: 6 EAVSGNIGGRHIVAAGQDINVRGXSLISDKGIVLKAGHDISTAHRYTNEYHESXX 65
+AV + G ++I+ +G+DI V G ++I+D +L A ++I + A R E ++
PspA: 1 235 QAVSGTLDGKEILLVSGRDITVGTSGNIADNHTLSAKNNIVLKAETRSRSEHNKKE 1294

Orf116: 66 XXXXXXXXXXXXXXXXXXXXNRTNIVHTGSLINGDTYVAGRIYQTSSTVSFE 125
 +*K + HT S++GSLNG++ AG Y QTGST+SSE+
 PspA: 1295 SGLMSGGIGFTAGSKDTQNRSEYSHTESVGLNHTLISLNGTSTISSPQ 1354
 Orf116: 126 GRNTYAKXIDVEFANRRVATDYAHTEQNGKLTVALNPRXXX---XXXXXXXKXKS 182
 G +++ I ++ A RRY+ + EOKG+TVA++VE
 PspA: 1335 GDVGLSSGKTSIDAAQNRYSOESHQYEQKQVTVASVPVNTVAGADVAVAVQVGRS 1414
 Orf116: 183 KNRKRYXXXXXXXXXXQSQYQATQHQFPA--PSSSAGQGNQYNSPSISYIXYEOKSN 240
 KN RV + + A P +AOGG ISVS+ YGEOK+ +
 PspA: 1415 KNSRVNMAAANAALNKGKVDQSGVALYNAANPKKAAGG-----ISVSYTYGEOKNTS 1466
 Orf116: 241 EOKRYTEAASQIIIGKQDTLAATSGSQSNINITGSDVYGHACTLLADNHRISQAK 300
 E + T* + I G G+ +L A+G+G+ S I ITGSDV G GT L A+N ++++++
 PspA: 1467 ESRIGTQVEKRTIGGKRVSLTASGKDSRILITIGSDVYGGKCTRLAKENAVQLEAR 1526
 Orf116: 301 QDSQSQSNKSSGMAAGYAKKIGRFRGITAKXXXXXXXSTTHRTHTVSTTGT 360
 Q E+S+NKS+G+NAQV I GI FG TA T +R+H+GS +T
 PspA: 1527 QTHOERSKNSAGFMAGVAMALNKGISFGTACANYGKVGNDETAYNSHISKDSQT 1586
 Orf116: 361 TIRSGGDTLKGVQLIKGQIADRLNLTIESVDTEYOSKQNGNVTYVYGFSAQS 420
 I SGGDT +KG QL GKG+ +LHIES+QDT ++ KO+N + QVTVYGFSS GS
 PspA: 1587 AIESGDTVIKGGOLKKGKGVVTAESLHIESLDTFAVFKGKQENVAQTGVGFVSQGS 1646
 Orf116: 421 YRQSRVKAADNASVTGSGSGLYAGEDGQIQVRNDTLKGLITTSQSAEDKRNLTFTATL 480
 Y +SK +D+ASV OSQI+AG DGY+I+V T L G + S DK RNL +T* +
 PspA: 1647 YNRSSSDVAASVNEQSGIFAGGGRFRIKRVNGKTLGVGAAYVSD---AKSKNLTLTSEI 1703
 Orf116: 481 TASDQMSHTEGNSFEIGGSEF 502
 DIONH+ + G+ G F
 PspA: 1704 WHKDIOHMAASALGSLGGF 1725

Based on homology with pspA, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 15

35 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 65>

```

1 ..ACAGCCGCA GCTCGGGCG CATACTGCC GCGGGCGCA CTTCCTTGC
51 CGACCGTAT TTGACCAAG CGGCGGAAA CTCGCTCG GCGGGCAAG
101 CGGCGTCAA CGCAGTGGC GGTGGCGCA TCGGCTATG AACTGGTGT
151 AGTGTGTG CTGTGGTGG TCGAATGTA GATTGAAAC ATAGCACT
201 GCATCGAAA GAATGGCGT TGGCGGAAA ATATCGGAA GCGCCGAGC
251 CGCAAGITCA AAAACGCGA GGCAGGAAA TATGCGGAA AGAAGCGGCA
301 ATGAGATCC GCAGGAGAT ATGCGTGGG TGCACAAAG TTCCCAACAC
351 GCGTATACG ACCAAAGCT CATATCCCTT ATCGAATCA

```

This corresponds to the amino acid sequence <SEQ ID 66; ORF118>:

```

1 ..TTGSLGGLA GGGTSLAAY LDKAENLGP AGKAVNALG GAATGATGG
51 SGGAVGANV DMNRDLHPR EMALADYAE ALKREYKNE GRKISQZAA
101 HRIRQLCVG WTKPRTAIR TKASPISE+

```

Computer analysis of this amino acid sequence reveals two putative transmembrane domains.

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 16

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 67>

```

1 ..CATGCCGTC TGAANAAGTC ACAATTATC AGACGGCATT TGTATTGCA
51 GTACATATAC AGATTCTCTA TACTATCCC AGACGGTGC GTGTGTGAG
101 ACACCCCCCA CGCTTCTAT TTG+PAAAGC TCCAACTAC CAAAGACATC
151 AACTGAGAC AGAT+ACT GCGTGAAGC AAATGGACT ATAAACGGA
201 AGGTTTACC GGAAGCTGAG CAGCATTTAT TGGCTGGCT GTTACGGTG
251 TTACTCGGGA CGGG+AGCC GGAGCCGAC TGGG+TTAA CGGGGGGCG
301 GCAAGCGGCA CGGATGCCG ATTGCGCTCG CTGGCCACC AGGCTTCGT
351 ATGCTCATC AACAA+AAAG GCAATATCG TACACCTCG TACCGAGCG
401 GCAAGAGCG CACCGTGAIA ANTCTAGTG TTGCGCTGCG TACCGAGCG
451 GTAGCGACA AATCTGTGC TTGGCCACTG AACATGTGA AGCTTACGA
501 GTGGATCAC AACCTAGCG TCACCTGCG CAATCGGGC AGCTTACGA
551 TGATTATAC CGCTGTCAAC GCGCGAGCG TGAAGAGCG TGTGAGACG
601 AATATCTTG CGCTTTGCT GAATATCGG CATGAGAGC CAGCCAGTAA
651 AATCAACAG TTGGATCAG ACTACATAC CCACAAGTT GCCCTAGCA
701 TAGGGGCTG TCGGTTGCG GCGCGAATA AGGCGAGCT TGCAGATGT
751 GCATATGCT CGCTGTGCG CGAATATGTC GGGGAGGCTT TGACAAAGG
801 CAAAAATCT GACCTTTGA CAGCTAAGA ACCGAGACG ATTGTGCT
851 ACAGCAACT GGTGCGCGT ACCGTAAAG GTCGTGCG CGGCATGTA
901 AATGGGCGG CGATTCGCG TAGAGTACG GTGAATAATA ATGACCTTG
951 GCACAAAGA

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF41>:

```

1 ..OCRLKSSQFY RRHLIKYIY RPIYCPKAC VAETPPACI LXOLQTTQV
51 NNNOYXLYAD KWDYKQEGIT GAGAILIATA VIVYFAGGA GAGALONGA
101 AANTDAFAS LASQASVSLI NNRKNGITL KELGSGSTY NLAVAVATG
151 VADKIGASAL NNVSQDWIN NLITANLNGS SALLNVAIVG GGLSKNLEA
201 NILALVNTA HGEAAKIKO LDQYITHKI AHAKAGCAA AANKKCDG
251 AIGAVERIY GEALTNKRP DTLTAKENQ ILNISKLVG TVSGVGGDV
301 NAAHNAEVA VKNNOJSDK+

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

1 ATGCAAGTAA ATATTGAT TCCCTATTA CTGCCCAAT GCGTCGTGC
51 TGAAGACAC CCCATCGATT GCTATTGGA ACAGTCCAA GTACACCAAG
101 AGCTCACTG GAACAGGTA CACTGGCGT ACGAAGTGA GACATATAA
151 CAGGAGAGCT TAAACGAGC CGAGACAGG ATTAATGCCC TGCGTTTAC
201 CGTGGTACT GCGGGCGCG GAGCGGAGC CCACTGAGC TTAAAGGCG
251 CGCGCGAGC GGCACCAAT GCGCAATTC CCGCTGCG CAGCCAGCT
301 TCGGTATCG TCATCAAGA CAAGGCAAT ATCGGTAGC CCGTAAAG
351 GCTGGGAGA AGCAGACAG TGAATAATC GATGGTTGC GTGGTACG
401 CAGGCTAGC CGACAAATC GGTCCTGCG CACTGACAA TGTCAAGAT
451 AAGCATGGA TCACAACT GACCTCAAC CTGCGGTAAT CGGCGATCG
501 CGCACTGAT TATACGCTG TCAAGCGCG CAGCTGAAA GACATCTGG
551 AAGCAATAT CTTGGCGCT TTGGTGATA CTGGCAATG AGAAGACGC
601 AGTAATATC AAGGTGGA TCAGCACTAC ATTACCCA AATATGCCA
651 TGCATAGCG GCGTGTGCG CTGCGGCGC GAATAGGCG AAGTGTAGG
701 ATGGTGCAT AGGTGCGCT GTGGCGAGA TAGTGGGGA GCGTTTACA
751 AAGCGAATA ATCTGACG TTGACACTT AAAGACGCG AACAGATTT
801 GGCATACAG AAATGTGTC CCGGTACGT AAGCGTGG GTGGGCGGG
851 ATGTAATCG GCGGCGAAT GCGGCTGAG AAGCGTGA AAATATCG
901 CTTAGCACA AAGAGGTAG AGAATTGAT AAGCAATGA CTGATCGCG
951 CAACAGAT ATCTCTAC TGTGAGAA AATATGTA AAAAGATAT
1001 AAATGTTC TGTAAAGA CTGTGCTT CATTGCAAT ATATGCGAT
1051 ATATCCGTA GTACTGAAT TACACATTC AGAAGCAAC ATTGATGCA
1101 TAGTAGAGC CTTCATCTT CTGGGAGCG AGCTTAAT GTTAAGATG
1151 ATGATGTA TAATTATTC AGCAATCTT ACACCAAC AGAATGGCT
1201 TTACAGTCT ATCAATTGA TACTGTCT AATCTTGC TTCAATGGG

```

1251 CAATACAAAG CTTTATCCG AATGATGTC GCACCAAGT TATACCTTA
1301 TTTCAGAGT TATCTTACG TTCTATCCAA TACCAGAGG CTTGTGAAA
1351 CAATATACG CTATTACTA TGTCAATAC CGGAGGCA CTGTTTCCA
1401 TACAACTTA AAGACATC TGGCAATGC TATGCTTTT AGTCAAAAC
1451 AGGCGATT AAGCGCCAT AGCGCAGCA ATTTATGCG AGAATTAAT
1501 TCACGAGAG GAGCGTAAA ATCTGAACC CAATATGTA TTGAAGGCT
1551 TACCGAAT TAATATGAG TTCTTACCT AGACGGCA GGTAACTTG
1601 ATGTGGAT TANGAAT TAAGTATA AACCTTTA TAATCTAAA
1651 AAATTTCTG ATGATAAT ACTTCAATG GCTCAAAATG CTGCTTCAA
1701 AGGATATCA AAGCTCTTA AATGCTGCA AATGAAGA ACTAATCAA
1751 TATCGGAAG AAAAATGTC ATCTAATCT CAGAACCTT TGCGGATC
1801 AATTAAGT CATATTTGA TGTAAATCA GGAAGAATTA CAACATCCA
1851 CCCAGAATAA

This corresponds to the amino acid sequence <SEQ ID 70: ORF41-1>:

15 1 MOVNIOIPVI LPRCVRAEDT PYACLYKQLQ VIKOVNNOV QLAYDWDYK
51 QGUTGAGAA IIALAVTVY AGAGAGALG VAGAATAATD AAFASLASOA
101 SVSLNNKGN IGMTLKEGR SSTVKNLHVA VATAVADKI GASALNNVSD
151 KOWNNLTVN LANAGSALI NTAVNGSLK DNLENTILAA LVNTANGENA
201 SKIKOLDORH ITHKIALTA GCRAAANKG KODQALGAA VGEIVGEALT
251 NGRKPDOLTA KERQILHATA KUVAGTVSGV VEGDVNNAAN AAEEVKNNO
301 LSKDEGREFT NEMTACAKN NPOLCRNNTV KTYONVADKR LAASIAICTD
351 LRSRETRTI RKHLIDSRG LHSMEAGLI GKODWYKLF SKSYTOADLA
401 LQSYHLNTAA KSWLQSGNTL PLSMNSDOG YLLISGVNPR FPIPIRGVFK
451 QNTETVNVKY PEGISFDNL KRHLANADGF SKQKQINGAR NRTNFMALN
501 SRGRVYKSET OTDIEGTETI KYEITPLDRT GKPDGGRFRI SIKTYVNPK
551 KFSDDKILQM AQNAAGSGYS KASKIAQNER TKSIERNV IQFSETFDGI
601 KFRSYFDVNT GRITNHPK

Computer analysis of this amino acid sequence predicts a transmembrane domain, and homology with an ORF from *N. meningitidis* (strain A) was also found.

30 ORF41 shows 92.8% identity over a 279aa overlap with an ORF (ORF41a) from strain A of *N. meningitidis*:

35 orf41.pep YRHLCKYIRFPYCPXACVAEDTPYACTLQJOVTKOVNNOVXLAYDKWDYKQEG
orf41a YLQLOVAKNNVNOVQVLAAYDRWDYKQEG
70 80 90 100 110 120 129
TGAGAAIIALAVTVVYTAGAGAGALGLANGAAAATAAFASLASQASVSLNNKNGT
orf41a TEGAAIIALAVTVVYTAGAGAGALGLANGAAAATAAFASLASQASVSLNNKNGT
130 140 150 160 170 180 189
LKLGRSSTVKLVAVATAGVADKIGASALNNVSDKQWNNLTVNLANAGSAAALNTAV
orf41a LKLGRSSTVKLVAVATAGVADKIGASALNNVSDKQWNNLTVNLANAGSAAALNTAV
190 200 210 220 230 240 249
NGGSLKONLEANIILAVNTANGAASKIKOLDORHLYTHKTAHAIAGCAAAANKGQD
orf41a NGGSLKONLEANIILAVNTANGAASKIKOLDORHLYTHKTAHAIAGCAAAANKGQD
160 170 180 190 200 210

250 260 270 280 290 300 310 320 330
orf41.pep GAI GAAGVEIVGEALTNKGNPDILTAKEREQILAYSLKLVAGTVSGVGVGDNAAANAAEV
orf41a GAI GAAGVEIVGEALTNKGNPDILTAKEREQILAYSLKLVAGTVSGVGVGDNAAANAAEV
310 320
orf41.pep AVNNQLSDRX
orf41a AVNNQLSDXREGREDFENENTACAKQXKPOLCRNNTVKKYQNVADKRLASIAICTDISR

A partial ORF41a nucleotide sequence <SEQ ID 71> is:

15 1 ..TATCTGAAC AGCTCCNAGT AGCGAAACAC ATCACTGGA ATCAGGTSCA
51 GTCTGCTTAC GACAGATGGG ACTACAACCA GAGGEGCTTA ACGRAGCAG
101 GTGCGGCGAT TATCGCAGTG GCGGTACCG NCGTCACTC AGCGCGAGG
151 ACGGAGCGG TATGGGATT AAAGGTGGG NCGCGCGCG CAACCGATGC
201 AGCATTCGCC TCTTTGGCCA GCCAGGCTT CGTATCGTTC ATCAACACA
251 AAGCGATGT CGGCAAAACC CTGAAGAGC TGGGCAAGG CAGCAGCGT
301 AAAATCTGG TGGTTCGCG CGCTACCGCA GCGTAGCGG ACMAATCGG
351 CGCTCGGCA CTGACNATG TCAGGATTA GCAGTGGATC AACACCTGA
401 CGGTCAACT AGCAATGCG GCGAGTGGG CACTGATTA TACCGCTTT
451 AAGCGGCGA GCCTGAAGAA CANTCTGGA GCGAATATC TACCGCTTT
501 GGTCAATACC GCGATGAGG AAGTACGCA TAAATCAAA CAGTGGATC
551 AGCACTACAT AGTCCACAG ATGCGCATG CCATAGCGG CTGTCGCGA
601 GCGGCGGGA ATAAGGCGAA GTGTGAGGAT GTGCGGATG GTGCGGCTT
651 GCGCGGATA GTGCGGAGG CTTTGACAAA CGGCAAAAT CTTGACACTT
701 TGACGCTAA AGAACGCGA CAGATTTGG CATAACGCA ACTGTTGCC
751 GGTACGGTA GCGGTGTGT CCGGCGGAT GTAAATGCG GCGGAGTGC
801 GCGTAGGTA GCGGTGAAA ATAATCAGT TAGCGACNAA GAGGTAGAG
851 AATTGTATA CAAATGACT GCATGCGCA ACAGATATN TCCTCACTG
901 TGCAGAAAA ATACTGTAA AAGTATCAA ATGTTGCTG ATAAAGACT
951 TGCTGCTCG ATTGCAAT GTACGATAT ATCCGTAGT ATGAATGTA
1001 GCAATCAG AAAACAAT TTGATGATA GTAGAGCTT TCATCATCT
1051 TGGGAGCAG GTCTAATTTG TAAGATGAT GAATGTATA ATATTTCAG
1101 CAATCTTAC ACCAAGCAG ATTTGGCTT ACAGCTTAT CATTTGATA
1151 CTGCTGCTAA ATCTGGCTT CANTGCGCA ATACAAAGC TTATCCGAA
1201 TGGATGCGG ACCAGGTTA TACACTTAT TCAGGATTA ATCTAGAT
1251 CATTCOAATA CAAAGAGGT TTGTAAACA AAATACACT ATTACTATG
1301 TCAATACCC GAAAGGATC AGTTTGTATA CAACTTANA AGCACTCG
1351 GCAATGCTG ATGCTTTAG TCAGACAGG GCGTTAAG GAGCCATAA
1401 CCGACCAAT NTATGGCAG AACTAAATC ACAGGAGGA NGNTAATAT
1451 CTGAACCCA NACTGATAT GAGGCAATTA CCCGAATTA ATAGGAT
1501 CCACTACTAG ACAGACAGG TAACTCTGAT GGTGATTA AGGAATTC
1551 AAGTATAAA ACTGTTATA ATCTTAAA NTTTNGAT GATAAATAC
1601 TCAATGCG TCANATCT GNTTCAAG GATATTCAAA AGCCTCTAA
1651 ATTTGCTCAA ATGAAGAAC TAAATCAATA TCGGAAGAA AAAATGAT
1701 TCAATCTCA GAACTTTG AGCAATCAA ATTAGANNN TATNTGATG
1751 TAAATACAG AGAATTACA ACATTCACC CAGATAA

This encodes a protein having the partial amino acid sequence <SEQ ID 72>:

50 1 YLQLOVAKN INNVQVLAAY DRWDYKQEG TELGAALIAL AVTVVTSAG
51 TGNVLGUNA XAAATDAFA SLASQASVSF INNRGVGKT LKELGRSSTV
101 KNLVVAATA GVADKIGASA LKNVSDQWI NNKTVNLANA GSAALNTAV
151 NGGSLKDXLE ANILAAVNT AHGAASKIK OLODHYIYKH IAHAIACAA
201 AAANKGQD GAIGAAGEI VEGALNKN PDILTAKERE QILAYSLKVA
251 GTVSVGVDG VNAANAAREV AVNNQLSDX EGREFNENT ACAKQNPOL
301 CRNTVTKYQ NVADKRLAAS IACTDISR EGRFENENT ACAKQNPOL
351 WEAGLKQD EYKLFKSY TOADLQSY HNTAKSWL QSDNTPLSE
401 WNSDQVTLI SGVNPFIPI PRGFVQNTP ITNVKYPEGI SGTNLRHL
451 ANADFSQD GIKGAHNTN XHAEINRGG YNKSETXDOI EGITRIKYEI

PCT/IB99/00103

Amino acids 25-619 of ORF41-1 were amplified as described above. Figure 6 shows plots of hydrophobicity, antigenic index, and AMPPH regions for ORF41-1.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 73>

This corresponds to the amino acid sequence <SEQ ID 74: ORF51>:

Further work revealed the complete nucleotide sequence <SEQ ID 75>:

30
1 ATCCAAAGAA TAAAGCAATC TATCGTTTT GTCTCTGCGG CAAATCTGCA
51 CCGAATATCA CAGGAGGAAT TTCCATCTCT GGGTCAAMC GCATGGCTGT
101 TTATCAATCG ATTTGCTAAC GTGTGGCTCT GTGGTGCAT ACCAAGCTGT
151 TATAAGCATC TGTGTGGTAT ATGCAAGAT ACCAAMAAAG GTTTTGGCA
201 AAGAGATGTT TATTATTTAA AAGCAATAA ATTGCTCTC ATCCGACAGC
251 TCGCTGGCAG CATTTGGAGG GTAGAGTTG TTTGTAACT TCCAGTGTGT
301 TGCGCTGTTT TACTAGTCTC AATCATTCAC TTGTAAATTT CTGTCAATGG
351 TATTAAATTA GTATGTGCGA AACCAAMAA TTTGATATTT CTGTCAATGG
401 ATACAAATTA GTCTCTTTT GGGTTTTGG CAGCAATCAT GGTGGGCTTCA
451 ACCATCAATCA TGTCTCCCAT ATTGTAAAT TTTTCTCTA GCGAACAAGA
501 AATTAATAAT CGATCTGCTA AATCAAGCA TCAATCTCAT TCTTTGGCGA
551 AAAATGTTCA ATTAATATGTC CTAAAGAGCC AGAATGGTGT ATTAATATAG
601 AGTAAATAGG GTTAAATAT TTATCTGCCC GATTCGTCTG TTAATGGATAT
651 GTATCTGGGA ATTCGGTTAA GCACTAAAT TACCACTAAT TTTTGAATAA
701 TGTTAAATTT TATGTCTTAA TTGGTATTTG CTTCAAAAT GGGGCAATTCG
751 GGTATTAATCA AACTTTAA

45 This corresponds to the amino acid sequence <SEQ ID 76; ORF51-1>.

1 MOEIMQSVIF VAAAILHGTT GMGFPMIGTT ALAFIMPLSK VVALVALPSL

51 LMSLLVCSN NKGFWQEV YLKYTKLLA IGSVGSILG VKLLILPVS
101 WLLLMHAIIT LYSVNGILN VCAKAKIQV VANNKMWLF GELAGIIGES
151 TRANSPIILI FLASETENK RAKSNILCY LLAKIVQIYM LRQVWLLNK
201 SEYGLIFLLS VLSVGLVYG IRLTRKISP FERMFLIFVL LVALKIGHS
251 GLIKL*

5

Computer analysis of this amino acid sequence reveals three putative transmembrane domains.

A corresponding ORF from strain A of *N. meningitidis* was also identified:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF51 shows 96.7% identity over a 150aa overlap with an ORF (ORF51a) from strain A of *N.*

10 *meningitidis*:

orf51a.pep 10 20 30
MALITLVSYNGILNVCARAKNIQVANNK
|||||

orf51a YKLAIGSVGSILGVKLLILPVSWLLLMHAIITLYSVNGILNVCARAKNIQVANNK
80 90 100 110 120 130

15

orf51a.pep 40 50 60 70 80 90
NHLVDFLXIIIGSTNAHSPILLIFLSETENNRIVKSNILCYLLAKIVQIYMLRQY
|||||

orf51a NHLVDFLAGIIGSTNAHSPILLIFLSETENNRIVKSNILCYLLAKIVQIYMLRQY
140 150 160 170 180 190

20

orf51a.pep 100 110 120 130 140 150
WLNKSEYGLIFLVSIGLVGIRLTKISNFEKLLIFVLVLLKIGHSGLIKL
|||||

25

orf51a WLNKSEYGLIFLVSIGLVGIRLTKISNFEKLLIFVLVLLKIGHSGLIKL
200 210 220 230 240 250

ORF51-1 and ORF51a show 99.2% identity in 255 aa overlap:

orf51a.pep HOETMOSIVFAAIIHGTGMPHIGTGTALAIHPLSKVVALVALPSLLMSLLVCSN
|||||

30

orf51-1 HOETMOSIVFAAIIHGTGMPHIGTGTALAIHPLSKVVALVALPSLLMSLLVCSN
|||||

orf51a.pep NKGFWQEVYLYTKYKLLAIGSVGSILGVKLLILPVSWLLLMHAIITLYSVNGILN
|||||

35

orf51-1 NKGFWQEVYLYTKYKLLAIGSVGSILGVKLLILPVSWLLLMHAIITLYSVNGILN
|||||

orf51a.pep VCAKAKNIQVANNKMWLFGLAGIIGSTNNSPILLIFLSETENNRIVKSNILCY
|||||

40

orf51-1 VCAKAKNIQVANNKMWLFGLAGIIGSTNNSPILLIFLSETENNRIVKSNILCY
|||||

orf51a.pep LLAKIVQIYMLRQVWLNKSEYGLIFLVSIGLVGIRLTKISPFFERMFLIFIVL
|||||

45

orf51-1 LLAKIVQIYMLRQVWLNKSEYGLIFLVSIGLVGIRLTKISPFFERMFLIFIVL
|||||

orf51a.pep LVALKIGHSGLIKL
|||||

50

orf51-1 LVALKIGHSGLIKL
|||||

The complete length ORF51a nucleotide sequence <SEQ ID 77> is:

1 ATGCAGAAA TAATGCATC TAICGTTTT GTTGCGCCG CAATACGCA
51 CGGAATTACA GCGATGGAT TCCGATGCT CGGTACARCC GCATGGCTT
101 TTATCATGCC ATTGCTAAG GTTGCTGCT TGGTGCAT ACCAGCGTG
151 TTAATGAGCT TGTGTTCT ATGCACCAAT ACACAAAGG GTTTTGCCA

201 AGAGATTGT TATTATTAA AACCTATAA ATTCTGCT ATCGCAGCG
251 TCGTTGCGAG CATTTGGGG GTGAAGTTC TTTTGATCT TCCAGTGCT
301 TGGCTGCTT TACTGATGG ATCATTACA TGTATATT GTGCAATGG
351 TATTTAAT GTATGCGAA AGCAAAAAA TAITCAATG TTTGCCAATA
401 ATAGAATAT GGTCTTTTT GGGTTTTTG CAGGCATCAT CGGCGGTCA
451 ACCAATGCCA TGTCTCCCAT ATTGTAATA TTTTGTCTA CGGAACACA
501 GAATAAAT CGATGCGAA AATCAAGCA TCTATGCTAT CTTTGGCAA
551 AATGTGCA AATATATG CTAGAGACC AGTATGGT ATTAATANG
601 AGTAATACG GTTAATATT TTACTGTC GTATGCTG TTTATGATT
651 GTATGTGGA ATTCGGTAA GGACTAAGT TAGCCAAAT TTTTATAAA
701 TGTAAATTT TATGTTTAA TCGGTATGG CTCTGAAT CCGGTATCA
751 GGTTAATCA ACITTTAA

5

10

This encodes a protein having amino acid sequence <SEQ ID 78>:

1 HOETMOSIVE VAAAILHGTI GMGFPMGLTI ALAFIMPLSK VVALVALPSL
51 LMSLLVCSN NKGFWQEV YLKYTKLLA IGSVGSILG VKLLILPVS
101 WLLLMHAIIT LYSVNGILN VCAKAKIQV VANNKMWLF GELAGIIGES
151 TRANSPIILI FLSETENK RAKSNILCY LLAKIVQIYM LRQVWLLNK
201 SEYGLIFLLS VLSVGLVYG IRLTRKISP FERMFLIFVL LVALKIGYS
251 GLIKL*

15

20 Based on this analysis, it is predicted that this protein from *N. meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 18

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 79>

1 ATGAGACATA TGAATAACA AATTATTATTA CTAGTATTA TAGTTTACA
51 TATAGCTTGG ATGATAATA ATATAGTGT TGGTATTT GTTTTCTAT
101 TTGATTTTT TCGTTTTTG TTTTGGCAA ACGTCTTCT TCGTGAAT
151 TTATATTTT TAGAAAAAA CATAAAAC AATTATTGT TTTTATGCG
201 GATTTCTAT ATATATGGA TGGTAATCA TATTAGTAG ATCCTGATT
251 AATTATTAA ATTTGACAT CAATAAAGG ACACAAAT ATCCTGATT
301 ACTGGGTGA TAAACACCA TGATAGTTAT AATTATGTT ATGACTCAA
351 TGGATATGCT AATTAAAG ATATCATAG ATATGTAGG GTAATTAGG
401 AACACCTTA TATTATGTA GTTGATCTG ATGTAATAA TAATCCATA
451 AGATTAGCT TGGTTGTGG TATTCATTA TATGCTCCAT GTGCCAATT
501 TATAAATTT GTCAGG..

25

30

35 This corresponds to the amino acid sequence <SEQ ID 80; ORF82>:

1 MHMKIQNYL LNFVULHIAL IVINIVGYE VLFDFEAF FFAVFLAWN
51 LFLFENIKN KLLFLPISI IIMVHISH INIKFYKEH OIKRONISSI
101 TGVIKPDSY NYVDSNGYA KLDNHRGR VIRETPYIDV VASDVNKS
151 RLSUVGSHS YAPCANFIRE VR..

40 Further work revealed the complete nucleotide sequence <SEQ ID 81>:

1 ATGAGACATA TGAATAACA AATTATTATTA CTAGTATTA TAGTTTACA
51 TATAGCTTGG ATGATAATA ATATAGTGT TGGTATTT GTTTTCTAT
101 TTGATTTTT TCGTTTTTG TTTTGGCAA ACGTCTTCT TCGTGAAT
151 TTATATTTT TAGAAAAAA CATAAAAC AATTATTGT TTTTATGCG
201 GATTTCTAT ATATATGGA TGGTAATCA TATTAGTAG ATCCTGATT
251 AATTATTAA ATTTGACAT CAATAAAGG ACACAAAT ATCCTGATT
301 ACTGGGTGA TAAACACCA TGATAGTTAT AATTATGTT ATGACTCAA
351 TGGATATGCT AATTAAAG ATATCATAG ATATGTAGG GTAATTAGG
401 AACACCTTA TATTATGTA GTTGATCTG ATGTAATAA TAATCCATA

45

This corresponds to the amino acid sequence <SEQ ID 82; ORF82-1>.

Computer analysis of this amino acid sequence reveals a predicted leader peptide.

A corresponding ORF from strain A of *N.meningitidis* was also identified:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF-82 shows 97.1% identity over a 172aa overlap with an ORF (ORF82a) from strain A of *N. meningitidis*.

25 or182.pep KLLELPISIIIMVHISMNIKFKFEHQIKQNISITGVIKRHSINYVVDNSGR

orf82 pep

KLLFLPILIIIMVHHSMTINKYKREHQKEONSSITGVAKRHDSNVYVSQNGZ
|||||
|||
orf82a

KLLFLPILIIIMVHHSMTINKYKREHQKEONSSITGVAKRHDSNVYVSQNGZ
|||||
|||

70 80 90 100 110 120

130 140 150 160 170

ORF82a and ORF82-1 show 99.2% identity in 242 aa overlap:

```

ORF82a      130      140      150      160      170      180
KLNKNRRYRYRIETREYEDVVAADKKNSRLSLVSGTNSAPCANCFKIFKFR
|||
ORF82-1      130      140      150      160      170      180
KLNKNRRYRYRIETREYEDVVAADKKNSRLSLVSGTNSAPCANCFKIFKFPKTYI

```

[illegible]

The complete length ORF82a nucleotide sequence <SEQ ID 83> is:

This encodes a protein having amino acid sequence <SEQ ID 84>

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>

1 .. ACCGCCAAG GCGTACGCT CTTCGCGTCT TTGGCGGAT TTGGGCGTAA
51 CGGCGCCAGC ATCATCTPAG CAGGCGGGGT CCGATATAGT GCGTTTGTGA
101 CAGCTTAAT TCTCTAGCC GAGGGGGGGT TGTAGAGAT GCGAGCGTGA
151 AAGGCAAAAG CCGTAAATGC AACGCGCGCT TGCATTTTAA CGGTCTGAA
201 TAGAGAAAT TTGCATTTCC TTITTAATTT CGGTTTTCAG ACAGCTAGCT
251 TCGCGCTGTA TTTCGCAAG ACCCATGCGC AAGAGTGGG CCTTAATTT
301 AATATTAAAA GCTTCGCGC GTGCCAGTTC CAGTTGGCC GCAATGTTT
351 GAGCGCAACA CACAGAGTCT TGGCGCTTGT CCGCTCTAAT CTTCGCTAAT
401 ACCGCTGCGA GCTTGCATAA TGGCAACTTG TACGCTTAAT GGTGGCAAA
451 AAGCAAGCCC GTTCGCAAGC GCGCGTAAT GGCATATGAT TGCAGTAA
501 TGGCCAGCAG TTTCACAGAA TTCAATCTG ACCTCTGAC GTTACAGCT
551 GA

This corresponds to the amino acid sequence <SEQ ID 86; ORF124>.

45 1 ..TENSIVULS EGGEGFICAT IMAGVOCHT AFSTILISA EGAVEILOQVA
51 AKKAVANITA CIETVLEKROD FOLEIFERFO TADRIYERFO SHADESVLD
101 RAKSVRACOF QERARIVSFO OGGLARVFO LVDRLOQARK CALVALVHR
151 SOABADRON GRNLPYIHO FHEITSREPD MSR

Computer analysis of this amino acid sequence predicts a transmembrane domain.

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

1	ATGACTGCTT	TTTGACAC	CTTAATTTC	GTAGCGAG	GCGGGTGT
5	AGAGTGAG	GCCTGTAG	CNAAAGCGT	CAATGCATCC	CCGCTTCTGA
10	TTTACAGT	CTTAGTAG	GACATTTTC	ATTTCCTTCG	
15	TTTCAGAGG	CTGATCTCG	CTGTATTTT	CGCAGAGCC	ATGCGGACAG
20	CGTGCSCCT	GACTTCATF	TTTTAGCTT	CCGCGGTGC	CAGTCCGAT
25	TCGGCGGAT	AGTTTGGAC	CNCAAGCC	AGGGCTTGG	CGTTCTGCG
30	CTCATCTT	TCGATAGCC	AGGCTGCTT	CGCAATGCC	CAGTGTAGC
35	TTATGTGTG	GCACAGCCG	CGACGCTGC	CGACAGTGC	GATATGTGCA
40	CTCGGTGCT	AGTTATTCG	CAGCAGTTTC	ACGAGATTC	TTCTCGACCT
45	CTTCAGTCC	CAGCTGTA			

This corresponds to the amino acid sequence <SEQ ID 88: ORF124-1>:

1 MTAFSTTLIS VAEGAVELQ AVRAKAVNAT ACICFTVLSK DIFDEIFEIFR
51 FQTADERLFF QSHADSVRL DIFEFSEFRAC QQCFANIVLS ROOQGRULVA
101 LHLVDORLIL RKCRULVALW RHUSQARADKR DNGNRILPVIR QQCFHEIHSRP
151 PDASR*

A corresponding ORF from strain A of *N.meningitidis* was also identified:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF124 shows 87.5% identity over a 152aa overlap with an ORF (ORF124a) from strain A of *N. meningitidis*:

oi124 . pep
 TPNSVLPSPGGFGRTGATINAGGVGHGTAFTSTLSVAEGNVLEQVRAKAVNTATA
 10 20 30 40 50 60
 oi124a
 HTAFTSTLSVAEGALVLEQVRAKAVNTATA

[illegible]

orf124 pep
QQLGLVALVILVDRLQLKCLVKLVALVWYHSQARAKPDNGINRFLPVTRQQFHEHSHRPD
| | | | | : : : | | | | | | | | | | | | | | | | | | |
| | | | | : : : | | | | | | | | | | | | | | | | | | |
QQLGLVALVFINDRLKLAKSLVAVLWYHQTRAKRDGDKGRFLPVTRQQFHEHSHRPD

orf124a
100 110 120 130 140 150
130 140 150 160 170 180

```
orf124.pep  ASRX
:
orf124a     VX
```

ORF124a and ORF124-l show 89.5% identity in 152 aa overlap:

or f124-1 pep	MTAFSTLISVAEGNVELQVRAKNAATACITFTVLSKQIDFELFIRFQTADERLFF	15
or f124a	MTAFSTLISVAEGALVELQVRAKNAATACITFTVLSKQIDFELFIRFQTADERLFF	
or f124-1 pep	RQSHADSVRLDIEFTSFRACOFOFARIVLSRQOGRLVALHVLVDLRLKRLVALMV	50
or f124a	RQSHADSVRLDIEFTSFRAFLRFAGVVLVSRCQOGRLVALHFLNDRLLRKSLVALMV	

orf124-1.pep R H S Q A R A D K R D G N R L P V I R Q Q F E I H S R P D A S R X
 || | : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
orf124a R H R Q T R A D K R D G N R L P V I R Q Q F E I H S R P D V X

The complete length ORF124a nucleotide sequence <SEQ ID 89> is:

5
0

1 ATGACCGCCT TTTCGACAC CTTATTTCG GTAGCCGAGG GCGCCTGTG
51 AGAGCTGCA GCCGTGATG CCNAGCCGT CATAACACC GCGCCTGCA
101 TTTTAGGTT CTGAGTAG GACATTTTG ATTTCCTTT TATTTCGGT
151 TTTCAGAGC CTGACTTCG CTGTTTTT CGCCNAGCC ATGCCNAGG
201 CTGCGCCTT GACTTCAT TTTTAGCTT CCGACGGCG CTGTCCAGT
251 TCGCGCGCT AGTTTAGC CAGCAGC AGGGTGGC CACTCTGAG
301 CTTATTITT TCAATGACG CTGCTGCTT CGCAAAAGC GTGTCTGAG
351 CTTCAGTGTG CGACACGCC ANCCCGTGG CGCAGCGG GATGATGCA
401 ATCGGTGGC AGTATTTCG CAGCAGTTTC ACAGANITCA TTCTCGACT
451 CCTGAGTTT GA

15 This encodes a protein having amino acid sequence <SEQ ID 90>:

1 HTAFSTTLLS VAEAGALVELQ AVMAKAVNTI AACIFTVLSK DIFDELFETFR
51 FQADSERLFF RQSHADGVRLL DEIFEFSTR LFQFAGVILS RQOQGLRIVA
101 LHEINDRLLL RKSRLVALMV RHROTADRKK DQGNRLPVIR QQFHEIHSRP
151 PDV*

ORF124-1 was amplified as described above. Figure 7 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF124-1.

Based on this analysis, it is predicted that this protein from *N.meningitidis*, and its epitopes, could be useful antigens for vaccines or diagnostics.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I - PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 38	Forward Reverse	CGCGATCCCATATG-TGCGCGAAAATTCGA CCCGCTCGAG-TTTCCCGCTTAAAGC	BamHI-NdeI XhoI
ORF 40	Forward Reverse	CGCGATCCCATATG-ACCGTGAAAGCCGCC CCCGCTCGAG-CCACTGATTAACGACAGA	BamHI-NdeI XhoI
ORF 41	Forward Reverse	CGCGATCCCATATG-TATTGAAACAGCTCCAAG CCCGCTCGAG-TTCTGGGTAAAGTTA	BamHI-NdeI XhoI
ORF 44	Forward Reverse	CGCGATCCCATATG-GGCACGCAACACCC CCCGCTCGAG-ACGTGGGAAACAGTCT	BamHI-NdeI XhoI
ORF 51	Forward Reverse	CGCGATCCCATATG-AAAAATATTCAGTAGTTGC CCCGCTCGAG-AAATTGATTTAAACCCG	BamHI-NdeI XhoI
ORF 52	Forward Reverse	CGCGATCCCATATG-TGCCAACCGCAATCCG CCCGCTCGAG-TTTTCCAGCTCCGGCA	BamHI-NdeI XhoI
ORF 56	Forward Reverse	GCAGATCCCATATG-GTATCGGAATATTACTCG CCCGCTCGAG-GGCTGCAGAAAGCTGG	BamHI-NdeI XhoI
ORF 69	Forward Reverse	CGCGATCCCATATG-CGACAGTGGTTGTTTT CCCGCTCGAG-ATACTTCCGTTTTTTTCAC	BamHI-NdeI XhoI
ORF 82	Forward Reverse	CGCGATCCGCTAGC-GTAAATTATTAATTTTAGAA CCCGCTCGAG-TTCCACTCAITGAAGTA	BamHI-NheI XhoI
ORF 114	Forward Reverse	CGCGATCCCATATG-AATAAGGTTTACATCGCAT CCCGCTCGAG-AATCGCTGCACCGGCT	BamHI-NheI XhoI
ORF 124	Forward Reverse	CGCGATCCCATATG-ACGCTTTTCGACA CCCGCTCGAG-GCGTGAAAGGTCAGGA	BamHI-NheI XhoI

TABLE II - Cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf38	+	+	+	His-fusion
orf40	+	+	+	His-fusion
orf41	+	n.d.	n.d.	
orf44	+	+	+	His-fusion
orf51	+	n.d.	n.d.	
orf52	+	n.d.	+	GST-fusion
orf56	+	n.d.	n.d.	
orf69	+	n.d.	n.d.	
orf82	+	n.d.	n.d.	
orf114	+	n.d.	+	GST-fusion
orf124	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, and 6.
2. A nucleic acid molecule which encodes a protein according to claim 1.
3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, and 5.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.
5. A protein having 50% or greater sequence identity to a protein according to claim 4.
6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, and 90.
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, and 89.
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, and 89.
11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.

12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8 to 11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 8 to 12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to *Neisseria* bacteria, particularly *Neisseria meningitidis*.

ABSTRACT

The invention provides proteins from *Neisseria meningitidis* (strains A & B), including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

FIG 1E

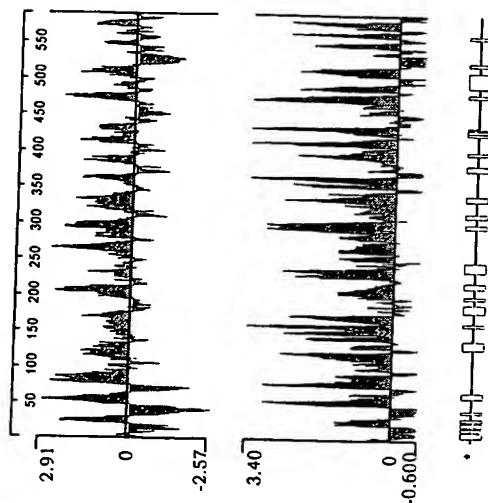


FIGURE 1



Fig. 1A



Fig. 1B

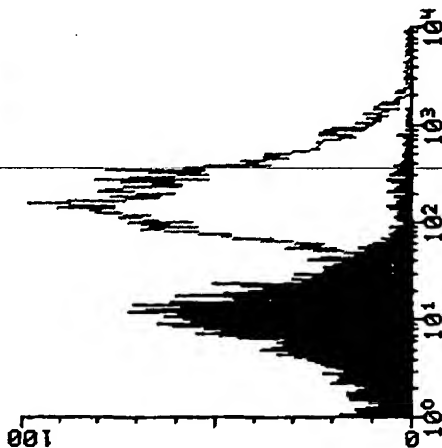


Fig. 1C

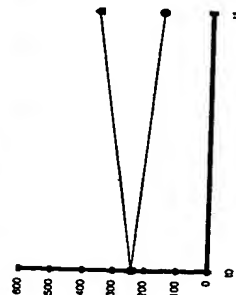


Fig. 1D

FIGURE 2

FIG. 2A



FIG. 2B

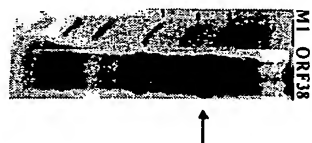


FIG. 2C



Fig. 2D

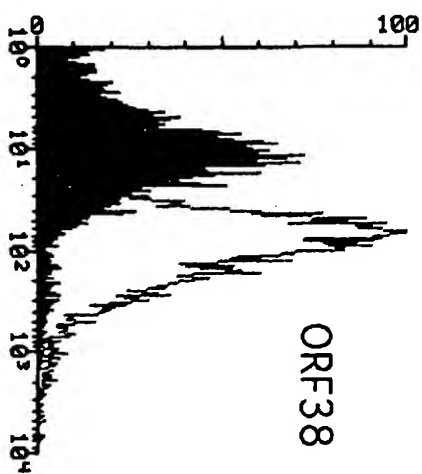


FIG 2E

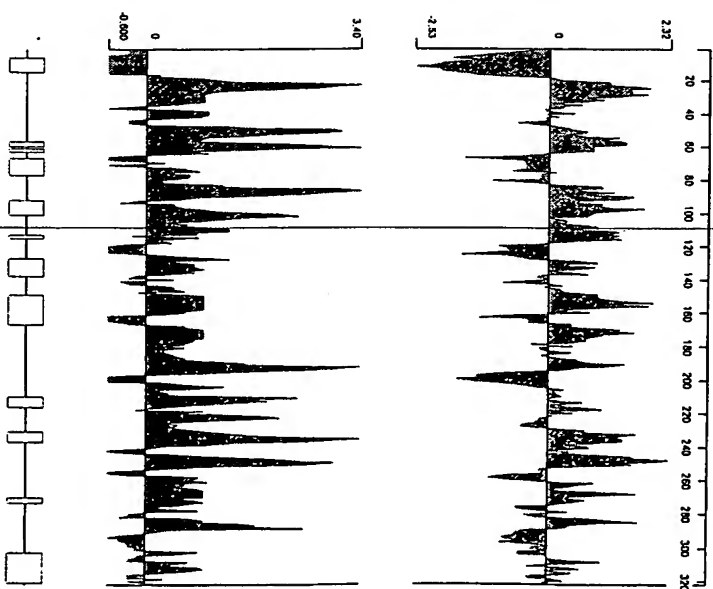


FIGURE 4

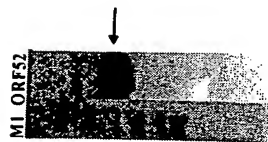


Fig. 4A

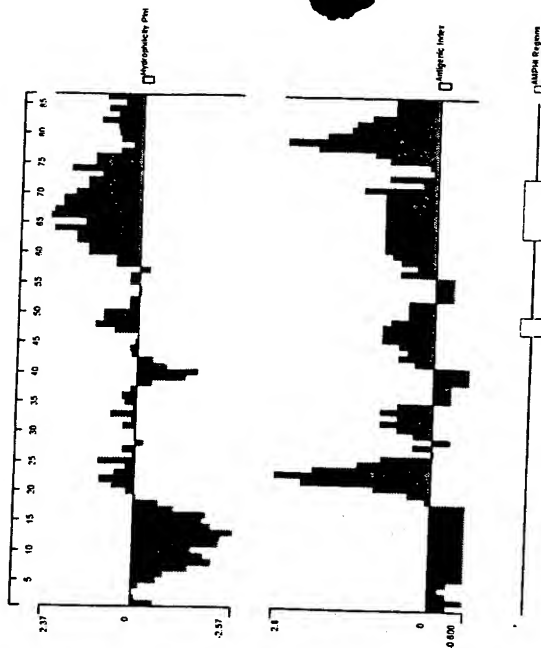


FIG. 4B

FIGURE 3



Fig. 3A

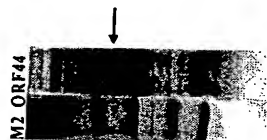


FIG. 3B

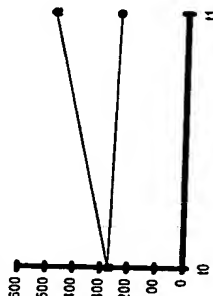


Fig. 3C

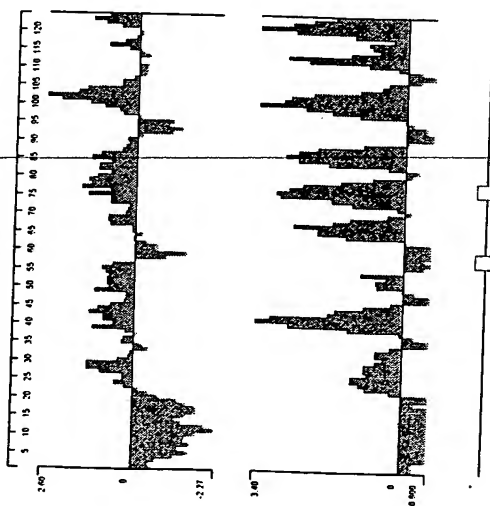


Fig. 3D

CLAIMS

1. A fragment of a protein disclosed in Annex 1, wherein the fragments comprise at least one antigenic determinant.
 2. The fragment of claim 1, having a length of 100 amino acids or less.
 - 5 3. The fragment of claim 1 or claim 2, having a length of 3 amino acids or greater.
 4. The fragment of any preceding claim, having an amino acid sequence disclosed in Table I.
 5. A polypeptide having 50% or greater sequence identity to the fragment of any preceding claim.
 - 10 6. A protein comprising one or more fragment of claim 1, claim 2 or claim 3, with the proviso that the protein is not one of the 45 complete protein sequences disclosed in Annex 1.
 7. An antibody which recognises the fragment according to any one of claims 1 to 4.
 8. A protein comprising a peptide sequence, wherein the peptide sequence is recognised by an antibody according to claim 7.
 - 15 9. Nucleic acid encoding the fragment of claim 1, claim 2 or claim 3, the polypeptide of claim 5, or the protein of claim 8.
 10. A composition comprising the fragment of claim 1, claim 2 or claim 3, the polypeptide of claim 5, the protein of claim 8, the antibody of claim 7, and/or the nucleic acid of claim 9, wherein the composition is a vaccine, a diagnostic reagent, or an immunogenic composition.
 11. The composition of claim 10 for use as a medicament
-
- 20 12. The use of the fragment of claim 1, claim 2 or claim 3, the polypeptide of claim 5, the protein of claim 8, the antibody of claim 7, and/or the nucleic acid of claim 9, in the manufacture of (i) a medicament for treating or preventing infection due to Neisserial bacteria (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria and/or (iii) a reagent which can raise antibodies
25 against Neisserial bacteria.
 13. A method of treating a patient, comprising administering to the patient a therapeutically effective amount of a composition according to claim 10.

